

Vihvelin on Frankfurt-Style Cases and the Actual-Sequence View

Carolina Sartorio

© Springer Science+Business Media Dordrecht 2014

Abstract This is a critical discussion of Vihvelin's recent book *Causes, Laws, and Free Will*. I discuss Vihvelin's ideas on Frankfurt-style cases and the actual-sequence view of freedom that is inspired by them.

Keywords Vihvelin · Frankfurt-style cases · actual-sequence views of freedom

Kadri Vihvelin's *Causes, Laws, and Free Will* (Vihvelin 2013) is an important defense of the classical compatibilist view concerning determinism and free will: the idea that determinism is compatible with free will, conceived as the ability to do otherwise. It is a fascinating attempt to draw on the metaphysics of dispositions and counterfactuals to develop the classical view and defend it on the face of objections. On Vihvelin's view, determinism is compatible with free will because causally determined agents can have the relevant (unexercised) powers to do otherwise, by virtue of having the relevant intrinsic dispositions and being placed in favorable, obstacle-free environments. Determinism is no threat to our free will because it doesn't undermine our power to do otherwise.

Vihvelin opposes an increasingly popular form of compatibilism that takes the ability to do otherwise to be irrelevant to the freedom and responsibility of agents: a type of view that had its origins in the so-called "Frankfurt-style examples" (Frankfurt 1969), and that has been developed in recent years by Fischer and Ravizza (1998), Fischer (2006), and others. Following common usage, I will refer to this other type of view as an *actual-sequence* view. This alludes to the fact that views of that kind focus only on the actual sequence of events, instead of on the availability of alternative possibilities.

I defend a version of the actual-sequence view elsewhere.¹ As an advocate of that kind of view, in this critical piece I focus on three interrelated themes in Vihvelin's book, all of

¹ See Sartorio (manuscript).

them connected with her discussion of Frankfurt-style examples and the view of freedom inspired by them. They are:

1. Vihvelin's claim that the difference between her view and the actual-sequence view is the difference between a metaphysical and a moral approach, and between a metaphysical and a moral form of compatibilism.
2. Vihvelin's argument that Frankfurt-style cases are not a threat to the idea that responsibility requires the ability to do otherwise.
3. Vihvelin's argument that other considerations connected with Frankfurt-style cases fail to cast doubt on the idea that responsibility requires the ability to do otherwise, and fail to motivate an actual-sequence view about freedom/responsibility.

In what follows, I take these up in turn.

Metaphysical and Moral Compatibilism

Vihvelin introduces the distinction between metaphysical and moral compatibilism in Chapter 1. She explains how the two forms of compatibilism can be seen as different responses to an incompatibilist argument that she takes to capture commonsense thinking. The following is a version of that argument as it applies to moral responsibility (p. 18):

1. If determinism is true, no one is ever able to do otherwise.
2. If no one is ever able to do otherwise, then no one is ever morally responsible.
3. Therefore if determinism is true, then no one is ever morally responsible.

Vihvelin calls premise 1 of this argument the *metaphysical* premise, and premise 2 the *moral* premise. Accordingly, she calls a form of compatibilism that attacks premise 1 a *metaphysical* compatibilism, and a form that attacks premise 2 a *moral* compatibilism. Thus, even if she herself is a compatibilist about determinism and *moral responsibility*, and not just about determinism and free will, she takes herself to be a *metaphysical* compatibilist about moral responsibility, in a way that sharply differentiates her view from the other forms of compatibilism, in particular, actual-sequence compatibilist views.

The distinction between metaphysical and moral compatibilism plays a central role in Vihvelin's book. These are her closing remarks:

It has been the main aim of this book to defend a compatibilist solution to the metaphysical problem of free will and determinism; I have argued that the intuitions that tempt us to believe that determinism deprives us of choice, ability to do otherwise, and alternatives should be resisted, for they are not, in the end, defensible. But even if I do not convince you of the details of my solution, I hope to have convinced you that the free will/determinism problem is *a problem in metaphysics and not, as current wisdom has it, a problem in moral philosophy*. (p. 238; my emphasis)

The "current wisdom" that Vihvelin alludes to here is the (increasingly popular) actual-sequence view or family of views.

Why does Vihvelin think that proponents of the actual-sequence view consider the problem of determinism and free will to be a problem in moral philosophy? She writes:

I will call these philosophers '**moral compatibilists**' for two reasons. First, *what* these philosophers defend is a moral thesis—the thesis that moral responsibility... is

compatible with determinism. Second, they defend this moral thesis... by arguments *within moral philosophy* rather than by arguments within metaphysics. (p. 18)

These claims by Vihvelin strike me as misplaced.

First, what those “moral”² compatibilists primarily aim to defend is not a moral thesis, but a metaphysical thesis: the thesis that a certain form of freedom or control is compatible with determinism. This is so even if they tend to identify the form of freedom or control as the one that is relevantly connected to moral responsibility (as “the sort of freedom or control that is relevant to responsibility”). Of course, they *also* want to defend a thesis about moral responsibility, the thesis that moral responsibility is compatible with determinism. But, note: so does Vihvelin (as I have mentioned, she wants to defend the compatibility of determinism and responsibility too).³ And (as is also the case with Vihvelin’s own view) this doesn’t mean that the thesis about responsibility is the thesis that moral compatibilists are *primarily* interested in defending.

The fact that one uses a moral notion to zero in on the concept that one takes to be relevant in a certain context doesn’t by itself turn the concept into a moral concept. It can still be a purely metaphysical concept, but the moral notion may be needed to pick it out and to differentiate it from other concepts. This should be pretty obvious, but it is worth emphasizing.

As an analogy, some philosophers have argued that there are multiple concepts of causation.⁴ In principle, it could be that only one of them, or only some of them, are relevant to the moral responsibility of agents, in the sense that they can ground their responsibility. For example, it could be that the concept of cause that is relevant to the sciences is not the concept of cause that grounds the moral responsibility of agents. In that case, it would make sense to use the notion of moral responsibility to identify the relevant concept of cause that helps ground it. This doesn’t make the relevant concept of cause a moral concept, or less of a metaphysical concept. Similarly, if there are different notions of control, as moral compatibilists are usually willing to allow,⁵ it seems perfectly reasonable to use the notion of moral responsibility to pick out the relevant concept (the concept of control that grounds responsibility), without this having any tendency to show that such a concept is a moral concept or less of a metaphysical concept.

Some moral compatibilists are very explicit about the fact that they are not primarily interested in defending a claim about moral responsibility, but one about freedom or control, understood as the partial grounds of responsibility. Fischer and Ravizza, in particular, explicitly note that they are interested in defending a view according to which determinism is compatible with a certain form of control (“guidance control”), which constitutes the freedom or control condition of responsibility. Moreover, they seem to conceive of that notion as primarily a metaphysical notion, since it is a form of control that concerns the specific causal mechanisms issuing in action and their properties (Fischer and Ravizza 1998; see especially Chapters 1 and 2).

² I’ll drop the scare quotes from now on. It should be clear that I think the label is not apt.

³ Vihvelin writes: “I’m not a pure metaphysical compatibilist; I work in the messy borderline between metaphysics and moral philosophy, and I want to defend the claim that we are morally responsible and that moral responsibility is compatible with determinism, *in addition to* defending the claim that we have free will and that free will is compatible with determinism.” (p. 19).

⁴ See, e.g., Hall (2004) and Hitchcock (2007).

⁵ In particular, Fischer and Ravizza (1998) distinguish “regulative” control from “guidance” control, and they take only guidance control to be relevant to moral responsibility. On their view, guidance control captures the freedom condition of responsibility (see below in the text).

At any rate, setting aside what moral compatibilists might have actually thought, the view that moral responsibility has a metaphysical component, roughly associated with a causal notion of control, is a very natural and plausible idea. Vihvelin seems to assume that any notion of freedom that comes apart from the ability to do otherwise is not a metaphysical notion. But this, it should be clear, is a highly dubious assumption.

So, Vihvelin's claim that moral compatibilism is primarily a moral thesis seems misguided. The same goes for her other claim: that the arguments used by moral compatibilists are arguments in moral philosophy. As we will see below, those arguments used by moral compatibilists tend to be based on certain intuitions about cases, in particular, Frankfurt-style cases. But, again, although the intuitions partly concern the moral responsibility of agents, they more fundamentally seem to concern the degree of causal control or freedom that those agents exhibit, which serves to *ground* their responsibility. So, the claim that those arguments are arguments in moral philosophy also seems misplaced.

To be clear: I think that Vihvelin's emphasis on the relevance of certain metaphysical debates and concepts to the problem of determinism and free will is highly commendable and important. I wholeheartedly agree with her that there is not enough emphasis on the relevance of metaphysical views and concepts in the current literature on free will, and that this is a gaping hole in the literature. (In fact, elsewhere I also argue that we should be drawing much more heavily on certain metaphysical concepts, in particular the concept of cause, in studying the problem of free will and in developing a compatibilist view of freedom.)⁶ What I contest is Vihvelin's claim that moral compatibilists don't do, or can't do, metaphysics. I think it is clear that they can, that they should, and that some of them in fact do.

Frankfurt-Style Counterexamples

Vihvelin's view, again, is the view that having free will amounts to being able to do otherwise, in the relevant sense. More specifically, in Chapter 1, she argues that the relevant ability to do otherwise is what she calls a "wide" ability. Wide abilities are contrasted with narrow abilities. Narrow abilities are abilities that we have just in virtue of our having certain intrinsic properties. Wide abilities, by contrast, are abilities that we have in virtue of our intrinsic properties *and* in virtue of being placed in certain surroundings. For example, I have the wide ability to leave the room when I have the relevant narrow ability (basically, the relevant physical and psychological capacities to walk out of the room) *and* when I have the means and opportunity to exercise that ability, in particular, when nothing external stands in the way (when there are no obstacles to my leaving the room, such as the door being locked from the outside). Vihvelin argues (rightly, to my mind) that the abilities that are relevant to free will cannot be the purely narrow abilities, since those are not the abilities that we seem to care about when we face "choice" situations—when we are deliberating to decide whether to do something or to do something else instead. As a result, throughout the book, she takes the relevant abilities to be wide abilities, not narrow abilities.

Frankfurt-style cases seem to challenge the idea that responsibility requires the ability to do otherwise, understood in this way, as a wide ability, not a narrow ability (the idea that responsibility requires the ability to do otherwise is typically known as the "principle of

⁶ In Sartorio (manuscript), I draw on certain metaphysical properties of causation to build a compatibilist actual-sequence view of freedom. See also Sartorio (2012).

alternative possibilities”). In a Frankfurt-style case, an agent, say, Jones, chooses to perform, and succeeds in performing, some action X. Given the way in which Jones comes to perform X (on the basis of his own reasons, upon freely deliberating about his options and his reasons, etc.), Jones appears to be morally responsible for X. However, there is another agent waiting in the wings, say, Black, a powerful and resourceful neuroscientist. Black secretly wants Jones to perform X, and has it in his power to prevent Jones from choosing to do anything but X. But, given that Jones chooses to perform X completely on his own, Black doesn’t need to intervene. So, it appears that Jones is responsible for choosing to do X and for doing X even if, given Black’s presence, he couldn’t have done otherwise (again, in the sense of wide abilities; that is, Jones lacks the relevant wide ability to do otherwise and to choose to do otherwise).

Vihvelin argues that Frankfurt-style cases fail to undermine the idea that responsibility requires the (wide) ability to do otherwise. She classifies Frankfurt-style cases into two main kinds: “bodyguard” cases and “preemptor” cases. In bodyguard cases, what would have triggered Black’s intervention is the beginning of any *action* by Jones (overt or mental) contrary to Black’s desires. Had Jones begun to do or to try to do anything other than choosing to X, Black would have intervened and would have prevented him from succeeding, say, by ensuring that he still chooses to X. In these cases, Vihvelin argues, Jones lacks the wide ability to make a different choice. For, even if Jones has the narrow ability to make a different choice (he has the relevant physical and psychological capacities), Black is an obstacle to the exercise of that narrow ability. However, Vihvelin argues, in these cases Jones retains the wide ability to at least *begin to choose otherwise*, where this, again, is a voluntary action. And this, Vihvelin argues, is enough to make him responsible for his choice. So Jones has the ability to do otherwise, in the relevant sense.

In preemptor cases, by contrast, Black’s intervention is triggered by some *earlier* and *non-actional* event, such as an involuntary sign by the agent (say, a blush or a twitch) that happens right before the agent is about to begin to decide to do something else, and that reliably indicates that the agent is about to begin to decide to do something else. Vihvelin’s diagnosis about these cases is very different. In these cases, she argues, Jones retains the ability to do otherwise in the relevant sense because, despite Black’s presence, he retains the wide ability to *make a different choice* (to choose to do other than X). This is because, first, he retains the narrow ability to make a different choice, and, second, he has the means and opportunity to exercise that ability; there are no external obstacles to his exercising it. In particular, Vihvelin argues that, given that Black never intervenes, Black *isn’t* an obstacle to the exercise of Jones’s narrow ability to make a different choice. Vihvelin compares this with an example involving a fair coin and non-agential abilities. If, on a certain occasion, someone with reliable predictive powers can predict that the coin will come up heads when tossed, he could have intervened before it’s tossed but didn’t because he wanted it to come up heads, the existence of the predictor/potential intervener doesn’t deprive the coin of the ability to come up tails: it’s still the case that the coin could have come up tails, in the relevant sense (it has “what it takes” to come up tails, and there are no obstacles to its coming up tails).

I think that both of Vihvelin’s arguments (the argument about bodyguard cases and the argument about preemptor cases) are problematic. I’ll discuss them in turn.⁷

⁷ Another potential criticism of Vihvelin’s treatment of Frankfurt-style cases is that her taxonomy doesn’t seem to be exhaustive, because it leaves out some scenarios that have been discussed in the literature (such as “trumping” and “blockage” cases). I won’t focus on this issue here.

Bodyguards

There seems to be an important gap in Vihvelin's argument about bodyguard cases. Her argument assumes that, given that in these cases Black's intervention is triggered by some *actional* event, such as the agent's beginning to choose to do otherwise, or to try to choose to do otherwise, the agent retains a robust enough ability to do otherwise to ground his responsibility. But in recent years some philosophers have argued against precisely this assumption. In particular, some philosophers have focused on Frankfurt-style cases where the agent retains an ability of this kind but, still, this isn't enough to ground his responsibility.

Consider, for example, a version of Pereboom's "Tax Evasion" case:

Tax Evasion: Joe decides to cheat on his taxes. His psychological profile is such that he couldn't have decided otherwise without first achieving a certain level of attentiveness to the relevant moral reasons. Jones is a libertarian free agent who can as a result of his voluntary activity make it the case that the relevant moral reasons occur to him with a certain force. Such attentiveness to moral reasons is causally necessary but *not* causally sufficient for his deciding not to cheat on his taxes: were Joe to achieve this attentiveness, he could still, with his libertarian free will, either choose to act on it or refrain from doing so. Now, to ensure that Joe chooses to evade taxes, Black has implanted a device that, were it to sense a moral reason of the relevant kind occurring with a certain force, would hijack his mental processes and ensure that he decide to cheat on his taxes. As it turns out, however, Black never has to intervene, because Joe never achieves the relevant level of attentiveness to moral reasons. (Pereboom 2001, pp. 18–19)

In this case, Black's intervention would be triggered by a mental action by Jones, namely, his bringing himself to achieve a certain attentiveness to moral reasons (which we are assuming he could have achieved as a result of voluntary activity). Thus this seems to fit the description of a bodyguard scenario, according to Vihvelin's taxonomy of Frankfurt-style cases.

But, then, Tax Evasion can be used to show that there is a gap in Vihvelin's argument about bodyguard cases. Pereboom argues that, although Joe has some alternative possibilities (he could have brought himself to achieve a certain level of attentiveness to moral reasons), they aren't sufficiently robust to ground his moral responsibility for tax evasion, even if (as we are assuming) they concern actional events under his control. This is because bringing himself to achieve that level of attentiveness to moral reasons is, by assumption, not sufficient for Joe to avoid responsibility, since it's compatible with his still deciding to evade taxes, and Joe is aware of all of this. And, if those alternative possibilities aren't sufficiently robust, Joe's responsibility for evading taxes *cannot* be grounded in his decision not to achieve the relevant level of attentiveness to moral reasons. So, Joe is responsible for evading taxes even if he lacks a robust ability to do otherwise. While Vihvelin might be right that an agent like Joe retains *some* wide (or non-narrow) abilities to do otherwise, he arguably still lacks the *relevant* wide abilities (those that are relevant to responsibility, or those that are responsibility-grounding).

At the very least, then, it seems that Vihvelin's argument about bodyguard cases is incomplete, since it doesn't address a specific subset of examples that have been the focus of some important debate in recent years.⁸

⁸ Franklin also hints at this gap in Vihvelin's argument in his review of her book (Franklin 2013).

Preemptors

Now let's examine Vihvelin's argument about preemptor cases. I should say at the outset that I had considerable trouble following this argument, and I am still not confident that I have fully grasped it (I will note some of the difficulties I ran into trying to reconstruct it along the way). But, to the extent that I was able to follow it, it seems to me that Vihvelin's argument about preemptor cases is also problematic.

Again, one difference between bodyguard and preemptor cases is supposed to be that, in preemptor cases, Black would have intervened earlier, *before* Jones could even begin to try to do otherwise, triggered by some reliable *non-actional* sign. So, imagine that Black would have intervened right before Jones began to make such an attempt, triggered by his blushing at that time. Vihvelin's view seems to be that in that case Jones retains *all* of the relevant abilities, including the ability to do otherwise, to choose to do otherwise, and to even begin to choose otherwise. She writes:

[I]nsofar as Black employs the method of the Preemptor, he makes *no difference whatsoever* to Jones's freedom—of will or action. For every alternative action—mental actions as well as overt actions—if Jones was able to perform that action before Black the preemptor came on the scene, Jones remains able to perform that action. (pp. 109–110)

Why does Vihvelin claim this, if Black's presence guarantees that Jones won't even begin to choose to do otherwise? She argues that it's because Black never actually intervenes: "To take away someone's freedom of the will, you must do more than Black does; you must actually mess with your victim's mind (brain) or body" (p. 112). Since Black never messes with Jones, in particular, he never messes with Jones's mind, Jones's freedom of the will remains intact.

Now, *why* is it that you can't take away someone's freedom of the will without actually intervening? On Vihvelin's view, there are some freedoms that concern the sphere of the will and that you can, in fact, take away without actually intervening. Vihvelin's discussion of bodyguard cases suggests this. As we have seen, her view about those cases, in particular, cases where Black's presence guarantees that Jones will make the relevant *choice*, is that Black deprives Jones of his ability to choose to do otherwise. This is a mental action, not an overt action, and in those cases Black can do this without actually intervening in any way. So Vihvelin's view can't be that, whenever Black remains a purely counterfactual intervener, his presence is completely irrelevant to the freedoms that concern the sphere of Jones's will.

What is Vihvelin's view, then? It seems to be the following. Factors that can undermine certain freedoms without actually intervening are what she calls "obstacles." Thus Jones's freedom of the will can only be affected by a purely counterfactual intervener like Black to the extent that Black is a real obstacle. For example, if you're in chains, your freedom of action is compromised because the chains work as an obstacle to your moving, even if you never try to move. Similarly, your freedom to make a different choice can be compromised in a bodyguard scenario if Black's presence is an obstacle to your choosing otherwise, even if you never try to make a different choice. (But recall that, even in that case, according to Vihvelin, you retain the ability to at least try to make or begin to make a different choice—that's why agents in bodyguard cases remain able to do otherwise, on her view.) Now, in a preemptor scenario, Black's presence guarantees that Jones won't *try* to make a different choice or *begin* to make a different choice. Thus Black can only be an obstacle if he is an obstacle to Jones's trying to make a different choice or to his beginning to make a different

choice. And, Vihvelin seems to think, nothing can be an obstacle of *this* kind without actually intervening.

But it is unclear what her argument for this is. She claims that whether something counts as an obstacle is something that is determined by the truth-value of certain counterfactuals. As Vihvelin points out, different counterfactuals apply to preemptor cases than to bodyguard cases (p. 98). The following counterfactual is true of bodyguard cases (but not of preemptor cases):

If Jones had begun to choose to do otherwise, or had tried to choose to do otherwise, this would have triggered Black's intervention, which would have prevented Jones from succeeding in making a different choice.

In contrast, the following counterfactual is true of preemptor cases (but not of bodyguard cases):

If Jones had shown some earlier sign that he was going to make a different choice, this earlier sign would have triggered Black's intervention, which would have prevented Jones from even beginning to make a different choice.

Vihvelin seems to think that the truth of the first counterfactual is sufficient to show that Black is an obstacle in the bodyguard case, but the truth of the second counterfactual is *not* sufficient to show that Black is an obstacle in the preemptor case. Why? I'm not totally sure. Perhaps this has to do with the fact that it seems odd to refer to something as an "obstacle" unless it's something that stands or would stand in the way of an agent *who is making a certain kind of attempt*. In the second case, what triggers Black's intervention is not an attempt by an agent, but an involuntary sign, so it seems odd to call Black an "obstacle" in that case.⁹

But, of course, the deep issue is not whether it's felicitous to call Black an "obstacle" in that kind of case, but whether Jones in fact retains the relevant wide ability to do otherwise. And the truth of the second counterfactual at least seems to suggest that Jones in fact lacks the relevant ability, given Black's presence.

Imagine, by analogy, this other scenario. A certain machine is equipped with a fail-safe mechanism that guarantees that the machine will never stop doing its job. The mechanism works by detecting early signs of malfunction: on any occasion, if the machine were to show any of those signs, it would intervene by immediately fixing the problem, before the machine even starts to fail. Imagine that the machine always works flawlessly, so it doesn't ever show any of the warning signs, and as a result the mechanism never intervenes. It seems that, under such circumstances, the machine couldn't even have begun to fail (it doesn't have the wide ability to begin to fail). This is, in fact, what the mechanism is there for, to make the machine completely fail-safe. Had the machine shown any of the early warning signs, the mechanism would have prevented it from beginning to fail. The truth of this counterfactual seems to support the claim that the machine lacks the relevant (wide) ability to begin to fail, in the same way that the truth of the counterfactual about bodyguard cases suggests that Jones lacks the relevant (wide) ability to make a different choice, when Black is a bodyguard. And the machine case seems analogous to the scenario where Black is a preemptor. So, if the machine lacks the relevant ability to begin to fail, it seems that

⁹ On p. 110, Vihvelin suggests that obstacles are things that do or would prevent attempts from succeeding, and she seems to take this as something like a conceptual truth. Perhaps, then, the implication is supposed to be that the concept of obstacle doesn't apply to attempts themselves. Thanks to Randy Clarke for discussion of this point.

Jones also lacks the relevant ability to begin to make a different choice, when Black is a preemptor.

An important question that arises for Vihvelin's view, then, is: why would it matter to Jones's having the relevant wide abilities if Black's intervention is triggered by an attempt by Jones or by some involuntary sign? At the very least, we would need an argument for the claim that this matters, and in the way Vihvelin suggests. Also, as we have seen, the argument cannot just be that the only way to affect an agent's freedoms concerning the sphere of his will is to mess with his mind. We know that there are other ways in which some of those freedoms can be affected. In particular, we know that, even on Vihvelin's own view, some of those freedoms can be affected simply by virtue of certain counterfactuals being true.

Now, there is another possible interpretation of Vihvelin's view about preemptor cases. This doesn't seem to be Vihvelin's own explicit or "official" view about those cases, but it is motivated by other things Vihvelin says later, when she discusses the relation between abilities, dispositions, and conditionals. So, let's examine this other possibility.

In section 6 of Chapter 6, Vihvelin discusses some common objections to the simple conditional analysis of abilities (the view that abilities can be analyzed in terms of simple conditionals, typically counterfactuals of some kind). On this type of view, "I could have done A" means something like "Had I chosen to do A, I would have done A." Chisholm (1964) famously raised the following objection to the simple conditional analysis, which has convinced many that the view is a nonstarter. Imagine that I have a pathological aversion that prevents me from being able to make a certain choice, say, I have a pathological aversion to red candy, because red candy reminds me of blood, so I can't bring myself to choose to take the red candy out of a bag, when it is offered to me.¹⁰ Then it seems that I don't have the ability to take the red candy from the bag. However, it could still be true that, *had* I (somehow) managed to choose to take the red candy, I *would* have taken the red candy from the bag. So, the conditional analysis entails that I have an ability (the ability to take the red candy) that I don't seem to have. I don't seem to have the ability to take the red candy because, more fundamentally, I don't have the ability to choose to take the red candy.¹¹

Interestingly, Vihvelin disagrees with this assessment of the case. She concedes that we have the intuition that I am unable to take the red candy in this case, but argues that the intuition is misguided. She argues for the view that, although I couldn't have *chosen* to take the red candy, I could have *taken* it. This is because this scenario is different from otherwise similar scenarios where I would have still had trouble doing something even after I had decided to do it. In the red candy case, as it is easy to assume, I would have easily managed to take the candy after having decided to take it (it only takes a second to do it, and it doesn't require any special skill). My body is not paralyzed; only my will is. Vihvelin argues that this suggests that I actually do have the ability to take the red candy; what I don't have is just the ability to choose to take it.

So, another way to interpret Vihvelin's view about preemptor cases is to take it to be the view that, whereas Jones *lacks the ability to begin to choose otherwise*, he retains the ability to choose otherwise (again, in the sense of wide abilities). This would be analogous to the way in which I lack the ability to choose to take the candy, but I retain the ability to

¹⁰ This is an example by Lehrer (1968).

¹¹ And, in these circumstances at least, my choosing to take the candy is necessary for my taking the candy. Clarke emphasizes this point in Clarke (2009: 329), as part of his criticism of Vihvelin's earlier views on this matter (Vihvelin 2004).

take the candy in the red candy scenario. Black's presence renders Jones unable to begin to choose otherwise because, had things been different in the relevant way (had he shown the blushing sign), Black would have intervened by preventing him to begin to make a different choice. But Jones retains the ability to make a different choice because, had he (somehow) begun to make a different choice, he would have succeeded in making a different choice. This is analogous to the way in which I would have succeeded in taking the red candy if I had, somehow, managed to choose to take it.

Again, this alternative view of preemptor cases doesn't seem to be Vihvelin's own considered view, but it might be well motivated if what she says about abilities more generally is true. So, it is still worth evaluating in its own right. On this view, then, Jones is able to do otherwise in the relevant sense because he retains the ability to make a different choice, even if he lacks the ability to begin to make a different choice.

Could this view work? I don't think so, but for a different reason this time. As I see it, the problem with this alternative interpretation of Vihvelin's view is that the kind of freedom that the agent in a Frankfurt-style case would have on this view (the relevant wide ability to do otherwise) doesn't seem robust enough to ground his responsibility. For, why would it matter if Jones can make a different choice (assuming he can, somehow, bring himself to try to make a difference choice), if he *can't*, more fundamentally, bring himself to try to make a different choice? Even if he has such an ability, it does not seem substantial enough to ground his responsibility.

It is noteworthy that, in the red candy case, Vihvelin refers to the ability that on her view the agent has as "the ability to take the candy *as the upshot of choosing to do so*" (p. 203; my emphasis). Note that here, too, one can grant that I have such an ability (especially when described in that way—which, at least to my ears, has some tendency to make it sound more like a "conditional" ability: the ability to do something *assuming* one can first choose to do so on the basis of reasons). But it's hard to shake the feeling that this ability is not robust enough to make me responsible for not taking the candy (especially if it is a mere conditional ability, but even if it's not a mere conditional ability). So, if I am right, this wouldn't only show that Vihvelin's argument about Frankfurt-style preemptor cases fails; it would also show that her argument that the conditional analysis has not been refuted by the red candy type of example fails too, assuming that the notion of ability that such an analysis tries to elucidate is a robust, responsibility-grounding notion. But I don't have the space to elaborate on this here.

I have distinguished two possible interpretations of Vihvelin's view about preemptor cases, and I have argued that they both face serious problems. On the first interpretation, Jones retains all the relevant abilities, including the ability to begin to choose otherwise. I argued that this view is not sufficiently motivated and that it's in fact hard to reconcile with Vihvelin's own claims about the connection between abilities and counterfactuals. On the second interpretation, Jones retains the ability to choose otherwise, but lacks the ability to begin to choose otherwise. I argued that this interpretation also has problems, because the ability that Jones has in such a case doesn't seem sufficiently robust to ground his responsibility. I conclude that Vihvelin's view about preemptor cases is also problematic.

The Supervenience Claim and the Actual-Sequence View

In section 7 of Chapter 4, Vihvelin discusses and criticizes an attempt by Frankfurt to appeal to the concept of *supervenience* (in particular, supervenience on actual sequences)

in order to cast doubt on the principle of alternative possibilities and motivate the actual-sequence view in its place. Frankfurt writes:

[I]f someone had no alternative to performing a certain action but did not perform it because he was unable to do otherwise, then he would have performed exactly the same action even if he *could* have done otherwise. The circumstances that made it impossible for him to do otherwise could have been subtracted from the situation without affecting what happened or why it happened in any way ... When a fact is in this way irrelevant to the problem of accounting for a person's action it seems quite gratuitous to assign it any weight in the assessment of his moral responsibility. (Frankfurt 1969, p. 174 of Watson 2003)

In this passage, Frankfurt seems to endorse a supervenience claim about responsibility. Vihvelin cashes out the supervenience claim in this way:

Facts about a person's moral responsibility *supervene* on the facts that causally explain why he did what he did; if two persons are alike with respect to all facts that causally explain why they act as they do, they are also alike with respect to their moral responsibility. (p. 118)

She then explains how the supervenience claim motivates the view that Jones, the agent in a Frankfurt-style case, is responsible (despite allegedly lacking the ability to do otherwise). By assumption, Black is not part of the causal explanation of why Jones acted as he did. Thus Jones in the Frankfurt-style scenario and Jones in an ordinary (non-Frankfurt-style) scenario are alike with respect to all facts that causally explain why they act as they do. But Jones is responsible in the ordinary scenario. So, it follows from the supervenience claim that Jones is responsible in the Frankfurt-style scenario too.

Vihvelin argues that this argument fails. She points out, the supervenience claim is ambiguous between a weak reading and a strong reading. The weak reading says that, when two persons are alike with respect to all the facts that causally explain why they act as they do, they are alike with respect to their moral responsibility *for something*: if one of them is responsible for something, then so is the other, where this is consistent with their being responsible for *different* things. The strong reading is committed to more than the weak reading: it says that, when two persons are alike with respect to all the facts that causally explain why they act as they do, they are alike with respect to their moral responsibility *for the same* things. (Here I am guessing Vihvelin means: for the same things, to the extent that they are relevantly connected with the action at issue; not, in particular, for things that they might have done years ago!) So, in particular, if one of them is responsible for the action at issue, then so is the other; if one of them is responsible for the choice to perform that action, then so is the other; etc.

Vihvelin first points out that weak supervenience doesn't give Frankfurt (or the actual-sequence theorist) what he needs. For, what the actual-sequence theorist wants to suggest is that Jones is responsible for *the very same* actions or choices that he would have been responsible for if Black hadn't been present. I agree completely with Vihvelin on this point. Interestingly, Fischer has recently argued for something like the weak reading over the strong reading of the supervenience claim (motivated, among other things, by the type of objection Vihvelin raises to the strong reading of the supervenience claim, which I'll discuss next; see Fischer 2013). I agree with Vihvelin that Fischer's retreat to the weak claim is not a good route for the actual-sequence theorist to take.¹²

¹² I discuss this in Sartorio (forthcoming).

So, that leaves the strong interpretation of the supervenience claim. Vihvelin's next step is to attack the claim on its strong interpretation. She offers two counterexamples to that claim. The first consists of the following pair of cases:

- *Smith1* Smith1 hears a child's cry for help just outside his office door. He decides to ignore it, and the child suffers some harm. Smith1 could have easily helped the child.
- *Smith2* Same story, except that, unbeknownst to Smith2, Black has locked Smith2's door; as a result, there is nothing Smith2 can do to prevent the child from being harmed.

Vihvelin claims that these two scenarios constitute a counterexample to the strong supervenience claim. In both cases, the agent doesn't come to the child's aid and the child is harmed; Smith1 and Smith2 are alike with respect to the facts that causally explain why they act as they do (say, in both cases, they decide not to help because they want to keep doing whatever it is that they're doing). However, they differ with respect to what they are responsible for: Smith1 is responsible for not aiding the child, and for the child's being harmed, but Smith2 is not (he is only responsible for making certain choices, for not trying to help the child, etc.).¹³

Vihvelin's second counterexample concerns Nagel's famous "resultant moral luck" cases involving two drunk drivers (Nagel 1979):

- *The lucky drunk driver* After driving drunk around town, the lucky driver arrives home safely because no one happens to run into his path while he was driving the car.
- *The unlucky drunk driver* While driving drunk around town, a child happens to run into the unlucky driver's path, and he kills the child.

Vihvelin claims that these scenarios also constitute counterexamples to the strong supervenience claim. For, what causally explains why they act as they do (driving drunk) may be exactly the same in both cases, but they don't appear to be responsible for the same things. (Or, Vihvelin claims, if one wants to say that they are responsible for all the same things because one rejects the possibility of resultant moral luck—the agents are only responsible for their decisions, and not for the outcomes of their decisions—then one's argument will rely on substantial, controversial assumptions about moral responsibility. Again, I agree with Vihvelin that this is not a route the actual-sequence theorist should take.)

I believe that the actual-sequence theorist has a good reply to these objections, though. I have argued for this elsewhere (especially in connection with the first type of alleged counterexample),¹⁴ but I'll rehash the crux of the reply here.

I think that we need to work with a slightly more precise formulation of the supervenience claim. In particular, the following schema will do for our purposes here:

(Strong Supervenience) Responsibility for X supervenes on the actual causal sequence issuing in X.

Here X can stand for anything for which agents could be responsible, including actions, decisions, omissions, outcomes, etc. But this formulation of the supervenience claim ties an agent's responsibility for something with the actual sequence issuing in that very same

¹³ As Vihvelin notes, others have raised similar objections to the supervenience claim; in particular, she credits van Inwagen (1983).

¹⁴ See Sartorio (2011 and manuscript). I actually think that the supervenience claim is a claim about the freedom condition of responsibility, not about responsibility *per se*, and only holds for responsibility when the epistemic conditions for responsibility are held fixed, but I won't get into this here, since Vihvelin's objections to the supervenience claim are objections of a different kind.

thing (not with the actual sequence issuing in some potentially different thing, as in Vihvelin's formulation). I think this is the formulation of the supervenience claim that makes the most sense, given the thought that the claim is trying to capture, namely, that one's responsibility for something depends exclusively on the way that *it* is brought about, or on *its* actual causal explanation. So, let's examine Vihvelin's alleged counterexamples, in the light of this improved formulation of the strong supervenience claim.

I'll start with Vihvelin's second counterexample (the two drunk drivers) because it's easier to see how the objection fails in that case, once the supervenience claim is formulated as Strong Supervenience. The two drunk drivers would be a counterexample to Strong Supervenience if there were something, X, such that the unlucky drunk driver is responsible for it and the lucky drunk driver is not, and such that the actual sequence issuing in X is the same in each case. But there is nothing that satisfies that description. In particular, if one takes X to be the child's death, the death doesn't even occur in the lucky driver scenario, so there simply is *no* actual sequence of events issuing in X in that case.

Now consider Vihvelin's first counterexample, Smith1 and Smith2. Again, this would be a counterexample to Strong Supervenience if there were something, X, such that Smith1 is responsible for it and Smith2 is not, and such that the actual sequence of events issuing in X is the same in each case. Here it might seem at first sight that something does satisfy that description; perhaps, even, that more than one thing does. Two good candidates seem to be: the omission to help the child (or the agent's not helping the child), and the child's being harmed. However, I will argue that the actual sequence of events is *not* in fact the same in these cases.

Consider the relevant outcome in the world: the child's being harmed. In both cases, the agent doesn't try to open the door. But, arguably, while Smith1's not trying to open the door is part of the causal explanation of the child's being harmed, a similar omission by Smith2 isn't. For, arguably, given that the door was locked from the outside, Smith2's not trying to open it didn't causally contribute to the child's being harmed. The same goes, I believe, for the omission to help the child. Smith1's not trying to open the door causally contributes to his not helping the child, but Smith2's not trying to open the door *doesn't* causally contribute to his not helping the child. So, the actual sequences are not in fact the same, perhaps despite initial appearances to the contrary.

Indeed, I would argue that the fact that the causal sequences are not the same explains why the agent is responsible for the relevant thing in one case but not in the other, and it does so in the most natural way possible. After all, why would the agent be responsible for the child's being harmed in one case but not in the other, if their causal connection to the harm were the same? Recall that their epistemic state, their intentions, etc. are by assumption the same in each case. So, if they both caused the harm, they would both have to be responsible for the harm, when they are not. This is not the place to develop this line of response in detail, but I believe this suggests that the actual sequences cannot be the same in these cases.¹⁵

I conclude that Vihvelin's attack on the supervenience claim fails. This is good news for the actual-sequence compatibilist. For, the supervenience claim is, arguably, one of the central elements of the actual-sequence view, if not *the* most central element of the view.¹⁶

¹⁵ I argue for this in more detail in Sartorio (manuscript).

¹⁶ I argue for this in Sartorio (manuscript).

Acknowledgments Thanks to Randy Clarke, Juan Comesaña, Michael McKenna, and Derk Pereboom for helpful comments and suggestions. Thanks also to Massimo Renzo for the invitation to contribute this critical review.

References

- Chisholm, R. (1964). "Human Freedom and the Self." *The University of Kansas Lindley Lecture*. Reprinted in G. Watson, ed., *Free Will* (Oxford University Press, second edition, 2003), pp. 26–37.
- Clarke, R. (2009) "Dispositions, Abilities to Act, and Free Will: The New Dispositionalism." *Mind* 118: 323–351.
- Fischer, J. and M. Ravizza (1998) *Responsibility and Control*. Cambridge: Cambridge University Press.
- Fischer, J. (2006) *My Way: Essays on Moral Responsibility*. New York: Oxford University Press.
- Fischer, J. (2013) "The Path of Life." Presidential address, given to the American Philosophical Association (Pacific Division).
- Frankfurt, H. (1969) "Alternate Possibilities and Moral Responsibility." *Journal of Philosophy* 66, 23: 829–839. Reprinted in G. Watson, ed., *Free Will* (Oxford University Press, second edition, 2003), pp. 167–176.
- Franklin, C. (2013) Review of Vihvelin's *Causes, Laws, and Free Will: Why Determinism Doesn't Matter*. *Notre Dame Philosophical Reviews*, <http://ndpr.nd.edu/news/43781-causes-laws-and-free-will-why-determinism-doesn-t-matter/>.
- Hall, N. (2004) "Two Concepts of Causation." In J. Collins, N. Hall, and L. Paul, eds., *Causation and Counterfactuals* (Cambridge, Mass: MIT Press), pp. 225–276.
- Hitchcock, C. (2007) "Three Concepts of Causation." *Philosophy Compass* 2, 3: 508–516.
- Lehrer, K. (1968). "Cans Without Ifs." *Analysis* 29: 29–32.
- Nagel, T. (1979) "Moral Luck." In *Mortal Questions* (New York: Cambridge University Press), pp. 24–38.
- Pereboom, D. (2001) *Living Without Free Will*. Cambridge: Cambridge University Press.
- Sartorio, C. (2011) "Actuality and Responsibility." *Mind* 120, 480: 1071–1097.
- Sartorio, C. (2012) "Causation and Freedom." *Journal of Philosophy* 109, 11: 629–651.
- Sartorio, C. (forthcoming) "Frankfurt-Style Examples." In M. Griffith, N. Levy, and K. Timpe, eds., *Routledge Companion to Free Will* (London: Routledge).
- Sartorio, C. (manuscript) *Causation and Free Will*.
- van Inwagen, P. (1983) *An Essay on Free Will*. New York: Oxford University Press.
- Vihvelin, K. (2004) "Free Will Demystified: A Dispositional Account." *Philosophical Topics* 32, 1–2: 427–450.
- Vihvelin, K. (2013) *Causes, Laws, & Free Will: Why Determinism Doesn't Matter*. New York: Oxford University Press.
- Watson, G., ed. (2003) *Free Will*. Oxford: Oxford University Press, second edition.