15

FRANKFURT-STYLE EXAMPLES

Carolina Sartorio

Introduction

We tend to think of ourselves as free and morally responsible agents. But what does acting freely (in the relevant sense) involve? According to the traditional view of freedom, it requires having the ability to select from a number of alternative possibilities that are open to us at different points in our lives. This idea is captured by the *Principle of Alternative Possibilities*:

> PAP: An agent acts freely (in a way that is relevant to his moral responsibility) only if he could have done otherwise (or had the ability to do otherwise).

There was a time when almost everyone believed in this principle. Incompatibilists argued that, if determinism is true, we never have the ability to do otherwise, and so determinism is incompatible with freedom and responsibility; compatibilists argued that determinism is compatible with having the ability to do otherwise, in the relevant sense, and so it is compatible with freedom and responsibility.

Harry Frankfurt's paper "Alternate Possibilities and Moral Responsibility" (Frankfurt 1969) radically changed the shape of the debate between compatibilists and incompatibilists. It argued against PAP, and thus against the traditional view of freedom. (Frankfurt's main concern was moral responsibility, but his argument can easily be reformulated in terms of the type of freedom that is relevant to responsibility, so this is how I will understand it here). Frankfurt's argument appealed to some hypothetical scenarios that have since then been called *Frankfurt-style examples*. Frankfurt-style examples have been the focus of much lively discussion in recent years, and the debate continues, as lively as ever, to this date.

Here, I'll focus on different aspects of the debate over Frankfurt-style examples, with a particular emphasis on the most recent developments, as well as on some issues that are still quite underexplored in the literature and that are likely to receive more attention in future years.

Frankfurt's Original Examples

Frankfurt argued that, although PAP seems initially plausible, its plausibility is a mere illusion. This illusion, he argued, can be explained away once a conceptual distinction

is made between two types of factors: (i) the factors that make an act by an agent inevitable, or that make it the case that the agent lacks the ability to do otherwise, and (ii) the factors that actually explain the agent's act. I'll refer to factors of the first kind as 'inevitability factors' and to factors of the second kind as 'explanatory factors.'

Frankfurt argued that the reason PAP has seemed initially plausible is that we naturally tend to conflate the two kinds of factors. When we think about scenarios where an agent lacks the ability to do otherwise, we naturally think of scenarios where the inevitability factors (those by virtue of which he cannot do otherwise on that occasion) are *also* the explanatory factors. For example, when an agent is coerced into doing something by an external threat, and thus doesn't act freely and is not responsible for what he does, the threat both makes his act unavoidable and explains his act. However, we can imagine scenarios where the inevitability factors *aren't* the explanatory factors. In some of those cases the agent seems to act freely and be responsible for his act, even though he lacks the ability to do otherwise.

Frankfurt illustrated this idea with examples. The most famous one involves a nefarious neuroscientist who plays the role of a counterfactual intervener. Here is a version of such a case:

> BLACK AND JONES: A neuroscientist, Black, wants Jones to perform a certain action. Black is prepared to go to considerable lengths to get his way, but he prefers to avoid showing his hand unnecessarily. So he waits until Jones is about to make up his mind what to do, and he does nothing unless it is clear to him (Black is an excellent judge of such things) that Jones is going to decide to do something other than what he wants him to do. If it were to become clear that Jones is going to decide to do something else, Black would take effective steps to ensure that Jones decides to do what he wants him to do, by directly manipulating the relevant processes in Jones's brain. As it turns out, Black never has to show his hand because Jones, for reasons of his own, decides to perform the very action Black wants him to perform.

In this case, Jones seems responsible for his choice because he made it completely on his own; Black never intervened. But it seems that, given Black's presence, Jones couldn't have made a different choice (Black wouldn't have let him).

So BLACK AND JONES seems to be a counterexample to PAP: it seems to show that being responsible doesn't require the ability to do otherwise. This is what we see, Frankfurt argues, when the inevitability factors come apart from the explanatory factors. Here, the inevitability factors are the facts concerning Black, his existence and intentions. Given that Black never intervenes, the inevitability factors are not part of the actual explanation of Jones's action. But, intuitively, the only factors that can be relevant to Jones's responsibility for his act are the factors that *actually* explain why he acted (the explanatory factors). The explanatory factors in this case are Jones's own reasons and process of deliberation. This is why Jones seems to be responsible even though he couldn't have done otherwise.

Frankfurt's argument has convinced many people, including many compatibilists and incompatibilists. But PAP continues to have advocates, who have responded to Frankfurt's argument in a number of different ways. In the next section, I explain the main moves, and how the dialectic has unfolded from there.

The Main Replies and the New Examples

A natural response on behalf of PAP is to say: in BLACK AND JONES, Jones isn't able to make a different choice, but he is able to at least *try* to make a different choice. Of course, had he tried, the neuroscientist wouldn't have let him make him a different choice, but it was open for him to at least try. So Jones still has alternatives, and those alternatives are enough to make him responsible for his choice. This style of reply is known as the 'flickers of freedom' reply (see van Inwagen 1983; Naylor 1984; McKenna 1997; Wyma 1997; Otsuka 1998; the label is from Fischer 1994: Chapter 7).

In turn, advocates of Frankfurt's argument counter-replied by offering Frankfurt-style scenarios where the alternatives that the agent has are not sufficiently 'robust' to ground his responsibility (see Fischer 1994, 1999; Pereboom 2001, 2014; McKenna 2003). Frankfurt himself imagined a scenario of this kind in a footnote to his original paper (Frankfurt 1969: footnote 3). He imagined that the basis for predicting what Jones was about to decide was an involuntary twitch of his face. Jones could have exhibited this sign or failed to exhibit it. But, obviously, Jones didn't have any control over this, since it's an involuntary movement of his face, so the fact that he has alternatives of this kind doesn't seem relevant to his responsibility. Frankfurt-style scenarios of this kind, where what would have triggered the neuroscientist's intervention is an early involuntary sign, are commonly known as 'prior-sign' examples.

The appeal to the concept of *robust alternatives* has played a key role in the literature on Frankfurt-style examples in recent years: much of the discussion has now shifted from the question whether it's possible to build Frankfurt-style examples without alternatives to whether it's possible to build Frankfurt-style examples without sufficiently robust alternatives. Why is this the right way to think about the debate over Frankfurt-style examples? Here is one way to see this. When we ask what our freedom consists in, we are not just interested in finding necessary and/or sufficient conditions for freedom; we are interested in finding conditions that can plausibly *ground* our freedom, in the sense that they are at least part of what makes us free, or of that by virtue of which we are free (the literature on the concept of grounding has itself grown a lot in recent years; for a recent overview, see Raven [2015]). Arguably, then, PAP should be interpreted as something like the view that freedom/responsibility are grounded in the availability of alternative possibilities (see Leon and Tognazzini 2010; Sartorio 2016). As a result, the relevant alternative possibilities are bound to be 'robust,' in the sense that they should be able to ground the agent's freedom. And the alternatives that the agent has in the prior-sign examples described above don't seem robust in this sense.

At this point, there is another natural reply on behalf of PAP, at least from the incompatibilist standpoint. It takes the form of a dilemma. Consider the relation between the involuntary sign and the agent's choice in prior-sign cases. There are two possibilities: either that relation is deterministic or it is indeterministic. So assume, first, that it's deterministic. Then the claim that Jones is responsible for his choice seems to beg the question against the incompatibilist, who believes that determinism is incompatible with responsibility. On the other hand, assume that the relation is indeterministic. Then there seems to be no reason to think that Jones's choice was unavoidable, for he could still have failed to make it, after having shown or not shown the relevant sign. Either way, there seems to be no counterexample to PAP (or, at least, to an incompatibilist version of it). This reply is commonly known as 'the dilemma defense' (see Widerker 1995; Ginet 1996; Kane 1996).

Again, advocates of Frankfurt's argument counter-replied in different ways. Some attacked the deterministic horn; others the indeterministic horn. Fischer (2010) attacked the deterministic horn by arguing that for a Frankfurt-style example to work we don't need to assume that the agent is responsible; all we need to assume is that, if he is not responsible, it's not because he wasn't able to do otherwise. As a result, Fischer argued, the argument based on prior-sign examples does not beg the question against the incompatibilist. Haji and McKenna (2004) developed a similar strategy of argumentation by distinguishing different dialectical contexts in which the debate about Frankfurt-style examples can be framed. Others attacked the indeterministic horn by offering different Frankfurt-style examples that are indeterministic in nature. The new examples are attempts to close off all (robust) alternatives without assuming determinism, while still preserving the intuition that the agent is responsible for what he does. Some of them involve other types of prior signs (Stump 1996, 2003; Haji 1998; Pereboom 2000, 2001, 2014; Hunt 2005); others don't involve any prior signs at all (Mele and Robb 1998, 2003; Hunt 2000; McKenna 2003). In Pereboom's case, for instance, the prior sign involves a condition that is *necessary but not sufficient* for the agent to make a different choice. The scientist knows the agent's psychological profile very well and he knows that, for the agent to choose otherwise, it is causally necessary that a certain moral reason occur to him with a certain force (perhaps as a result of his voluntary activity, of his forcing himself to consider that reason). But, if he were to make such a moral reason occur to him at any point, as far as the agent is aware he could still choose to ignore it. Unbeknownst to the agent, however, the scientist has implanted a device in the agent's brain to ensure that he makes a certain morally reproachable choice. The device would be triggered only if it sensed that the moral reason occurred to the agent with the specified force. As a matter of fact, however, that moral reason never occurs to the agent, so the scientist never intervenes. The agent appears to be responsible for his choice, but Pereboom argues that he lacks robust alternative possibilities. (Although he could have forced himself to consider the relevant moral reason, Pereboom argues that this isn't a robust enough alternative because, as far as he knows, he wouldn't have escaped responsibility for his choice simply by doing that, since he believed that he could have decided to ignore that moral reason).

The literature contains several other defenses of PAP from the threat of Frankfurt-style cases. The 'timing' defense urges us to pay close attention to time and argues that, although the agent in certain Frankfurt-style cases may be responsible for making the relevant choice *right then* (at the precise time at which he made it, which he could have avoided doing), he is not responsible for making the choice *simpliciter*, at some time or other (see Ginet 1996, 2002; Franklin 2011; Palmer 2011). The 'agent-causal' defense (Rowe 1991; O'Connor 2000) interprets PAP in terms of agent-causation instead of ordinary event-causation, that is, as the claim that responsibility for a choice that one agent-causes requires the power not to agent-cause it. Frankfurt-style cases are not counterexamples to this claim, for an agent in a Frankfurt-style scenario retains the power not to agent-cause his choice: if the neuroscientist had intervened, he would have failed to agent-cause his choice. (Sartorio [2012] discusses a similar reply for an event-causal version of PAP.) In turn, Alvarez (2009) and Steward (2008, 2009) have offered an 'action-theoretic' defense of PAP. Their defense is based on the claim that something is not a genuine *action* unless the agent could have refrained from performing it. If this claim is true, it follows that Frankfurt-style cases are not counterexamples to PAP (for, if it's really the case that the agent couldn't have refrained from performing it, then it's not a genuine action, and thus it falls outside of PAP's scope). There are also "dispositionalist"

defenses of PAP, which consist in arguing that agents in Frankfurt-style cases have the ability to do otherwise because they have all the relevant dispositions (see Smith 2003; Vihvelin 2004, 2013; Fara 2008). By appeal to the recent literature on dispositions, these solutions claim that the reason it might have seemed that agents in Frankfurt-style cases don't have the ability to do otherwise is that the relevant dispositions are 'masked' or 'finked,' given the presence of the neuroscientist (for a discussion of these ideas, see Clarke [2009]). Nelkin (2011) develops a similar strategy in terms of certain interference-free abilities that agents can retain even when placed in Frankfurt-style scenarios. Finally, there are some defenses of PAP that attempt to undermine the intuitions about Frankfurt-style cases by appeal to considerations about the nature of blameworthiness. Widerker's "W-defense" attacks the claim that an agent such as Jones is blameworthy by posing the question: "If he is blameworthy, then what should he have done *instead*?" (Widerker 2003). The fact that he couldn't have done anything different seems to suggest that he can't really be blameworthy. The debate over these issues tends to be intertwined with the debate over the plausibility of the ought-implies-can principle (the claim that a moral obligation to do something requires the ability to do that thing). For example, Fischer (2006) argues that the W-defense fails because Frankfurt-style cases also work as counterexamples to the ought-implies-can principle: Jones couldn't have done otherwise, but he still ought to have refrained from doing what he did. (For other views on this issue, see Copp 2003; McKenna 2008; Capes 2010; Moya 2011).

The Retreat from the Counterexample Strategy

As we have seen, Frankfurt's examples generated a lively debate that went far beyond Frankfurt's original argument. But, for the most part, the literature on Frankfurt-style scenarios has assumed that the examples are successful if and only if they constitute clear *counterexamples* to PAP. However, Frankfurt himself has recently challenged this assumption. He has argued that, even if the examples didn't succeed as direct counterexamples to PAP, they could still succeed in challenging the traditional view of freedom because they could help undermine PAP's *intuitive appeal* (Frankfurt 2003; see also Zagzebski 2000; McKenna 2008).

Recall the way in which Frankfurt originally introduced the examples. He did it by first drawing a distinction between two kinds of factors: inevitability factors and explanatory factors. He then explained how, although we tend to conflate them, the two types of factors can come apart, and he used Frankfurt-style examples to illustrate this. Once it becomes clear that inevitability factors and explanatory factors can come apart in this way, Frankfurt suggested, PAP loses much of its initial appeal. For then we can appreciate more clearly the fact that all that really matters to our freedom is the actual explanation of our behavior, and not whether we are able to do otherwise. As a result, even if it were not dialectically possible to offer any persuasive Frankfurt-style scenarios without (robust) alternatives, there would still be some reason for thinking that having alternatives is not relevant to acting freely and being responsible. So Frankfurt-style examples can play an important role in an argument against the traditional view of freedom without constituting direct counterexamples to PAP. This is a significantly different way of looking at the dialectic.

To illustrate this point further, consider the following analogy. Another important metaphysical concept that philosophers have struggled to give an account of is the concept of causation. Different accounts have been offered of that concept. A quite natural

one appeals to the concept of lawful sufficiency (i.e., logical sufficiency, given the laws of nature), or, if the laws are indeterministic, the more inclusive concept of probability-raising. A traditional 'regularity' view of causation is an account of that kind; roughly (assuming determinism) it says that something is a cause just in case it's a lawfully sufficient condition for the effect (or an indispensable part of some lawfully sufficient condition, as in Mackie [1965]). *Preemption* scenarios appear to be counterexamples to this idea. In a preemption scenario, there are two processes that are lawfully sufficient for an effect but only one of them (the preemptor) is actually efficacious; the other one (the preempted process) remains causally inert, as a mere backup. For example, if Suzy and Billy both throw rocks at a window and Suzy's rock gets there first, the process involving Suzy's rock preempts the process involving Billy's rock. Preemption scenarios are a problem for regularity views because the preempted process is lawfully sufficient for the effect without being a cause. So, in this case, preemption scenarios play the role that Frankfurt-style examples play in the debate over PAP. (They are also particularly apt in this context, for note that Frankfurt-style scenarios *themselves* are a type of preemption scenario, one where the causal result is a choice by an agent, the preempted process is the process involving the neuroscientist, and the preempting process is the agent's own deliberation).

Now, while these are not knockdown arguments against the regularity view of causation (advocates of the view, such as Mackie himself, claim that they can successfully deal with the problem), one could argue that preemption scenarios are still helpful in a deeper and more general kind of way. For they can open our eyes to the idea that lawful sufficiency (or probability-raising) is not *what really matters* to being a cause (at least as far as a *productive* notion of causation is concerned); what really matters is the existence of a physical connection, or of a continuous process, or of some other relation of a similar kind between cause and effect. One could argue, moreover, that the reason we couldn't see this clearly before is that, in ordinary cases, lawful sufficiency and this other relation tend to go hand in hand, whereas we can see this much more clearly in preemption cases, where the two come apart.

Similarly, then, perhaps Frankfurt-style cases can be helpful, *partly thanks to their artificiality*, because they can help us see more clearly a distinction that tends to be muddled in ordinary cases where agents are unable to do otherwise: the distinction between inevitability and explanatory factors. As a result, they can help us zero in on what really matters to freedom.

The Motivation for an Alternative View of Freedom

I have distinguished two different roles that Frankfurt-style examples can play: the role of acting as direct counterexamples to PAP, and the role of helping disarm PAP's initial plausibility. Both of these are *negative* roles, in that they are attempts to undermine a certain view of freedom. But there is also a *positive* role that Frankfurt-style examples can play, one that is related to the second negative role but also importantly different from it. This third role tends to be even more underemphasized than the second negative role, but it is also extremely important. It is the role of motivating a different view of freedom, one that is not at all based on the availability of alternative possibilities, but only on the relevant *actual sequences* or actual explanations of behavior.

Frankfurt's own view is that acting of one's own free will requires acting from a will that one wishes to have, that is, acting from first-order desires that suitably mesh with higher-order desires (see Frankfurt 1971; see also Watson [1975] for a development of a

non-hierarchical mesh view of freedom). But in recent years, Frankfurt-style scenarios have inspired and motivated some views of freedom that emphasize other aspects of the actual explanation of behavior, in particular, the *reasons-responsiveness* of agents or of the relevant mechanisms, and that have come to be known as 'actual-sequence views.' (The first comprehensive development of such a view is Fischer and Ravizza [1998]; see also Fischer 2006, 2012; McKenna 2008, 2013; Sartorio 2011, 2016). Again, these are actual-sequence views partly because they do not take the freedom of an agent to depend on the existence of alternative possibilities accessible to the agent. But the issue of how it is that Frankfurt-style cases motivate the new views, or what kinds of principles about freedom they support, *besides* the mere rejection of the alternative- possibilities requirement—in other words, the *positive* insights about freedom motivated by the examples—has not yet received enough attention in the literature.

Although Frankfurt's main goal in his original paper was to argue against PAP, he also suggested that Frankfurt-style examples motivate a different view of freedom, one uncommitted to PAP. Frankfurt suggested that reflection on the examples motivates the following principle about what is relevant to responsibility:

> RELEVANCE: The explanatory factors (those factors that actually explain the agent's behavior), and *only* those factors, are relevant to the agent's responsibility for his behavior.
>
> (Frankfurt 1969)

Again, Frankfurt thought that Frankfurt-style examples can help us see how the inevitability factors and the explanatory factors can come apart, and thus they can help us see the irrelevance of the inevitability factors and the relevance of the explanatory factors. So, at the same time that they can be used to undermine PAP, they can be used to motivate a different view of freedom, one according to which acting freely is just a matter of having a certain kind of actual causal history or actual explanation.

Now, RELEVANCE has been forcefully contested. What's more, the main actual-sequence views currently on offer don't seem to respect it. Here are a couple of apparent counterexamples to it:

> SQUEAKY BUTTON: I love squeaky sounds. I know that pushing a certain button would result in a squeaky sound, so I push the button to hear the sound. I also know that pushing the squeaky button will result in a remote village being wiped out, but that's not what motivates me to push the button (I'm not that evil!).
>
> (From Sartorio 2011, 2016. This style of objection was originally pressed by Widerker [2000, 2003], by appeal to similar examples. See also Palmer 2014.)
>
> SHARKS: A child is drowning in a pond nearby. I think I could easily rescue him, but instead of trying to save him I keep walking along the shore, unmoved by the situation, and the child dies. Unbeknownst to me, there were hungry sharks in the water. Had I tried to save the child, they would have attacked me, and I wouldn't have been able to rescue him.
>
> (From Fischer and Ravizza 1998. This style of objection was first pressed by van Inwagen [1983] and Ginet [1996], by appeal to similar examples.)

185

In SQUEAKY BUTTON, I am clearly blameworthy for causing the village's destruction. My belief that the village would be wiped out as a result of my pressing the button is clearly relevant to my responsibility (since, had I not been aware of the consequences of pushing the button, I wouldn't have been responsible for causing the village's destruction). However, that belief is not part of the actual explanation of my behavior, for I didn't push the button motivated by that belief (I only pushed it because I wanted to hear the squeaky sound). Thus, a factor that doesn't help explain my behavior can still be relevant to my responsibility for that behavior; in other words, RELEVANCE seems false. In turn, in SHARKS, I am clearly *not* blameworthy for not saving the child (of course I could be blamed for not *trying* to save him, but it seems wrong to blame me from not saving him, given that I couldn't have saved him). And the sharks' presence is relevant to my lack of responsibility (for, if the sharks had not been present, then I would have been responsible for not saving the child). However, the sharks are not part of the actual explanation of my behavior (they were not the reason I didn't jump in; I wasn't even aware of their presence!). Thus here, too, a factor that doesn't help explain my behavior can be relevant to my responsibility for that behavior.

So these examples suggest that RELEVANCE is false. If Frankfurt-style examples support a new view of freedom that is based exclusively on actual sequences, it doesn't seem to be one that entails RELEVANCE. Frankfurt responds to some of these objections in Frankfurt (2003) (I discuss some of his replies below). More recently, and motivated by the threat posed by the counterexamples, Fischer has proposed a weaker formulation of the principle, one that is designed to avoid at least some of them:

> RELEVANCE*: If a fact is irrelevant to the explanation of the person's action, then it is irrelevant to the issue of the person's "moral responsibility at all," i.e. to the issue of whether the agent is morally responsible *to at least some degree for at least something*.
>
> (Fischer 2013)

Note that this principle clearly avoids the problem with SHARKS (although it is less clear that it helps with SQUEAKY BUTTON). For the presence of the sharks *is* irrelevant to the issue of the agent's responsibility "at all": even when the sharks are present, I am still responsible *for something*, such as for not trying to save the child.

On the downside, however, RELEVANCE* seems too weak as an attempt to capture the insight behind Frankfurt-style cases. For Frankfurt-style cases don't only seem to motivate the idea that our responsibility *for something or other* (our responsibility "at all," in Fischer's terminology) depends exclusively on factors that actually explain how we act; they also, more strongly, seem to motivate the idea that responsibility *for something in particular* depends exclusively on the actual explanation of that very thing. For example, my responsibility for an action or a choice depends exclusively on how I came to perform the action or make the choice. And, similarly, my responsibility for an outcome such as the child's death depends exclusively on the actual causal history of the death (in particular, on my involvement in it).

Another way of formulating the view inspired by Frankfurt-style cases that has been explored in the literature, and that is also suggested by some passages in Frankfurt's paper, takes the form of a supervenience principle. Here is a possible formulation of that principle:

> SUPERVENIENCE: An agent's responsibility (freedom) with respect to X supervenes on the actual sequence issuing in X.

SUPERVENIENCE states that there cannot be a difference in responsibility (or freedom) without a difference in the actual sequence. In other words, whenever the actual sequence is the same, the agents are equally responsible (or free). Fischer once explicitly embraced SUPERVENIENCE, at least when X is an action by an agent (see Fischer 1987). (It is unclear what relationship Fischer would say exists between that principle and the principle discussed above, RELEVANCE*; unfortunately, Fischer doesn't discuss that connection explicitly). Sartorio (2011, 2016) embraces a generalized version of SUPERVENIENCE.

Like Fischer's principle RELEVANCE*, SUPERVENIENCE is weaker than RELEVANCE. For there could be 'extrinsic' factors (factors that are external to the actual sequence or causal history of an act) that still determine what the causal history is, or its actual composition, and thus, whether the act is free and whether the agent is responsible. In that case, SUPERVENIENCE could be true while RELEVANCE is false (extrinsic factors can be relevant to freedom even if freedom supervenes on actual sequences; see Sartorio 2011, 2016). Unlike Fischer's principle RELEVANCE*, however, SUPERVENIENCE doesn't seem *too* weak, or removed from the original insight offered by Frankfurt cases, for it still ties the agent's responsibility or freedom for something with the actual explanation of that same thing.

Do examples such as SQUEAKY BUTTON and SHARKS threaten SUPERVENIENCE in the same way they threaten RELEVANCE? Arguably, SHARKS is no longer a counterexample, for the causal history of the child's death is presumably different when the sharks are present and when they are not; in particular, when they are not present, my failure to jump in is a cause of the child's death, but not so when they are present. This fits well with the idea that I can only be responsible for the child's death, and thus for not saving the child, if I am causally connected to his death in the relevant way. A solution of this kind relies on a causal interpretation of actual sequences. Sartorio develops an actual-sequence view based on a causal interpretation of actual sequences, and argues that SUPERVENIENCE can accommodate examples like SHARKS in Sartorio (2016).

Conversely, SQUEAKY BUTTON seems to raise different types of issues from those raised by SHARKS. Compare the SQUEAKY BUTTON scenario with a similar scenario where I'm unaware of the destruction I'm causing by pushing the button. The causal history of my behavior and of the destruction seems to be the same in the two cases (in both cases I push the button because I want to hear the squeaky sound, and the village is destroyed as a result); however, there is a difference in my responsibility: I'm responsible for the destruction in one case but not in the other.

Now, here it might be important to bring to bear the distinction between freedom, or the freedom component of responsibility, and responsibility itself. When Frankfurt suggested the RELEVANCE principle, he seemed to have responsibility in mind, and many others who have followed his lead do too. But responsibility has an epistemic component too. Roughly, in order to be responsible one must be aware of the moral significance of what one is doing. And there is no obvious reason to think that what goes for the freedom component of responsibility should also go for the epistemic component. So a promising option (at least one that deserves further exploration) is to restrict SUPERVENIENCE or similar principles to the freedom component of responsibility, and to claim that freedom supervenes on actual sequences, although perhaps responsibility doesn't, due to its epistemic component (see Sartorio 2016). Otherwise put, responsibility only supervenes on actual sequences when the agent's epistemic state is

held fixed. This avoids the problem with SQUEAKY BUTTON, for the responsible and non-responsible agents are not in the same epistemic state.

When Frankfurt replied to the objections to the RELEVANCE principle, he suggested other possible ways of addressing the problem (see Frankfurt 2003; see also McKenna 2008). He suggested that examples like SQUEAKY BUTTON are not counterexamples to RELEVANCE, when that principle is properly understood, because the *full* explanation of the agent's act in that case includes the sort of act that he thought he was performing. Thus, if I am aware of the fact that I am acting wrongly by pressing the button, then a full explanation of what I do will include my awareness of my wrongdoing, even if it is not a cause of my act. Perhaps Frankfurt would say something similar about how the examples bear on SUPERVENIENCE. Perhaps he'd say that SUPERVENIENCE should be understood as the claim that an agent's responsibility for an act supervenes on the 'full' explanation of the act, in this broader sense.

This brings me to one final central issue. When advocates of the actual-sequence view say that freedom is just a function of actual sequences or actual explanations, what exactly should we take actual sequences or actual explanations to *be*? Initially, at least, it seems that a very natural interpretation is to take actual sequences to be just actual causal histories (this idea is developed in Sartorio 2016). As we have seen, Frankfurt himself doesn't seem to understand actual sequences in this way, for he takes the actually explanatory factors to include factors that go beyond the causes of the action. Fischer and Ravizza, in turn, seems to interpret actual sequences as including some modal properties of sequences that are not actually explanatory or causally efficacious in themselves, such as the 'reasons-responsiveness' of sequences (see Fischer and Ravizza 1998; see also McKenna 2013). Again, how we interpret actual sequences is tied to the issue of what we take the real insight behind Frankfurt-style examples to be. This is an issue that is still quite underexplored in the literature, and that deserves more attention than it has so far received.

Conclusions

The literature on Frankfurt-style examples is now huge, and rightly so, but it is likely to keep growing in new directions. As we have seen, the examples can play a variety of related roles in the debate about the nature of freedom and responsibility. First, they can be regarded as direct counterexamples to PAP, the principle that crystallizes the traditional view of freedom. Most of the debate about Frankfurt-style cases has focused on this role. But there are two other main roles for Frankfurt-style cases to play, which seem at least as important, if not more important. A second negative role is the role of supporting the argument against the traditional view of freedom more indirectly, by helping to undermine its initial plausibility, instead of by providing a direct counterexample to it. And a third and positive role is the role of motivating a different view of freedom, one according to which freedom is only a matter of the actual sequence having certain features, or of our acts being caused in the right kind of way.

Acknowledgments

Bibliography

Alvarez, M. (2009) "Actions, Thought-Experiments and the 'Principle of Alternate Possibilities,'" *Australasian Journal of Philosophy* 87: 61–81.

Capes, J. (2010) "The W-Defense," *Philosophical Studies* 150: 61–77.

Clarke, R. (2009) "Dispositions, Abilities to Act, and Free Will: The New Dispositionalism," *Mind* 118: 323–51.

Copp, D. (2003) "'Ought' Implies 'Can,' Blameworthiness, and the Principle of Alternate Possibilities," in D. Widerker and M. McKenna (eds), *Moral Responsibility and Alternative Possibilities*. Aldershot: Ashgate, pp. 265–300.

Fara, M. (2008) "Masked Abilities and Compatibilism," *Mind* 117: 843–65.

Fischer, J. (1987) "Responsiveness and Moral Responsibility," in F. Schoeman (ed.), *Responsibility, Character, and the Emotions: New Essays in Moral Psychology*. Cambridge: Cambridge University Press, pp. 81–106.

Fischer, J. (1994) *The Metaphysics of Free Will*. Oxford: Blackwell.

Fischer, J. (1999) "Recent Work on Moral Responsibility," *Ethics* 110: 93–139.

Fischer, J. (2003) "Responsibility and Alternative Possibilities," in D. Widerker and M. McKenna (eds), *Moral Responsibility and Alternative Possibilities*. Aldershot: Ashgate, pp. 27–52.

Fischer, J. (2006) *My Way: Essays on Moral Responsibility*. New York: Oxford University Press.

Fischer, J. (2010) "The Frankfurt Cases: The Moral of the Stories," *The Philosophical Review* 119: 315–36.

Fischer, J. (2012) *Deep Control: Essays on Free Will and Value*. New York: Oxford University Press.

Fischer, J. (2013) "The Path of Life," Presidential address, given to the American Philosophical Association (Pacific Division). Proceedings and Addresses of the American Philosophical Association 87: 60–77.

Fischer, J. and Ravizza, M. (1998) *Responsibility and Control: An Essay on Moral Responsibility*. Cambridge: Cambridge University Press.

Frankfurt, H. (1969) "Alternate Possibilities and Moral Responsibility," *Journal of Philosophy* 66: 829–39.

Frankfurt, H. (1971) "Freedom of the Will and the Concept of a Person," *Journal of Philosophy* 68: 5–20.

Frankfurt, H. (2003) "Some Thoughts Concerning PAP," in D. Widerker and M. McKenna (eds), *Moral Responsibility and Alternative Possibilities*. Aldershot: Ashgate, pp. 339–46.

Franklin, C. (2011) "Neo-Frankfurtians and Buffer Cases: The New Challenge to the Principle of Alternative Possibilities," *Philosophical Studies* 152: 189–207.

Ginet, C. (1996) "In Defense of the Principle of Alternative Possibilities: Why I Don't Find Frankfurt's Argument Convincing," *Philosophical Perspectives* 10: 403–17.

Ginet, C. (2002) "Review of *Living Without Free Will*," *Journal of Ethics* 6: 305–9.

Haji, I. (1998) *Moral Appraisability*. New York: Oxford University Press.

Haji, I. and McKenna, M. (2004) "Dialectical Delicacies in the Debate about Freedom and Alternative Possibilities," *Journal of Philosophy* 101: 299–314.

Hunt, D. (2000) "Moral Responsibility and Unavoidable Action," *Philosophical Studies* 97: 195–227.

Hunt, D. (2005) "Moral Responsibility and Buffered Alternatives," *Midwest Studies in Philosophy* 29: 126–45.

Kane, R. (1996) *The Significance of Free Will*. Oxford: Oxford University Press.

Leon, F. and Tognazzini, N. (2010) "Why Frankfurt-Examples Don't Need to Succeed to Succeed," *Philosophy and Phenomenological Research* 80: 551–65.

McKenna, M. (1997) "Alternative Possibilities and the Failure of the Counterexample Strategy," *Journal of Social Philosophy* 28: 71–85.

McKenna, M. (2003) "Robustness, Control, and the Demand for Morally Significant Alternatives," in D. Widerker and M. McKenna (eds), *Moral Responsibility and Alternative Possibilities*. Aldershot: Ashgate, pp. 201–18.

McKenna, M. (2008) "Frankfurt's Argument against Alternative Possibilities: Looking Beyond the Examples," *Noûs* 42: 770–93.

McKenna, M. (2013) "Reasons-Responsiveness, Agents and Mechanisms" in D. Shoemaker (ed.), *Oxford Studies in Agency and Responsibility*, Vol. 1. New York: Oxford University Press, pp. 151–83.

Mackie, J. (1965) "Causes and Conditions," *American Philosophical Quarterly* 2: 245–64.

Mele, A. and Robb, D. (1998) "Rescuing Frankfurt-Style Cases," *The Philosophical Review* 107: 97–112.

Mele, A. and Robb, D. (2003) "Bbs, Magnets and Seesaws: The Metaphysics of Frankfurt-style Cases" in D. Widerker and M. McKenna (eds), *Moral Responsibility and Alternative Possibilities*. Aldershot: Ashgate, pp. 127–38.

Moya, C. (2011) "On the Very Idea of a Robust Alternative," *Crítica* 43: 3–26.

Naylor, M. (1984) "Frankfurt on the Principle of Alternate Possibilities," *Philosophical Studies* 46: 249–58.

Nelkin, D. (2011) *Making Sense of Freedom and Responsibility*. New York: Oxford University Press.

O'Connor, T. (2000) *Persons and Causes: The Metaphysics of Free Will*. New York: Oxford University Press.

Otsuka, M. (1998) "Incompatibilism and the Avoidability of Blame," *Ethics* 108: 685–701.

Palmer, D. (2011) "Pereboom on the Frankfurt Cases," *Philosophical Studies* 153: 261–72.

Palmer, D. (2014) "Deterministic Frankfurt Cases," *Synthese* 191: 3847–64.

Pereboom, D. (2000) "Alternate Possibilities and Causal Histories," *Philosophical Perspectives* 14: 119–38.

Pereboom, D. (2001) *Living Without Free Will*. Cambridge: Cambridge University Press.

Pereboom, D. (2014) *Free Will, Agency, and Meaning in Life*. Oxford: Oxford University Press.

Raven, M. (2015) "Ground," *Philosophy Compass* 10: 322–33.

Rowe, W. (1991) *Thomas Reid on Freedom and Morality*. Ithaca: Cornell University Press.

Sartorio, C. (2011) "Actuality and Responsibility," *Mind* 120: 1071–97.

Sartorio, C. (2012) "Causation and Freedom," *The Journal of Philosophy* 109: 629–51.

Sartorio, C. (2016) *Causation and Free Will*. Oxford: Oxford University Press.

Smith, M. (2003) "Rational Capacities, or: How to Distinguish Recklessness, Weakness, and Compulsion," in S. Stroud and C. Tappolet (eds), *Weakness of Will and Practical Irrationality*. Oxford: Clarendon Press, pp. 17–38.

Steward, H. (2008) "Moral Responsibility and the Irrelevance of Physics: Fischer's Semi-compatibilism vs. Anti-fundamentalism," *Journal of Ethics* 12: 129–45.

Steward, H. (2009) "Fairness, Agency and the Flicker of Freedom," *Noûs* 43: 64–93.

Stump, E. (1996) "Libertarian Freedom and the Principle of Alternative Possibilities," in D. Howard-Snyder and J. Jordan (eds), *Faith, Freedom, and Rationality: Essays in the Philosophy of Religion*. Lanham: Rowman and Littlefield, pp. 73–88.

Stump, E. (2003) "Moral Responsibility without Alternative Possibilities," in D. Widerker and M. McKenna (eds), *Moral Responsibility and Alternative Possibilities*. Aldershot: Ashgate, pp. 139–58.

van Inwagen, P. (1983) *An Essay on Free Will*. Oxford: Clarendon Press.

Vihvelin, K. (2004) "Free Will Demystified: A Dispositional Account," *Philosophical Topics* 32: 427–50.

Vihvelin, K. (2013) *Causes, Laws, and Free Will*. New York: Oxford University Press.

Watson, G. (1975) "Free Agency," *Journal of Philosophy* 72: 205–20.

Widerker, D. (1995) "Libertarianism and Frankfurt's Attack on the Principle of Alternative Possibilities," *The Philosophical Review* 104: 247–61.

Widerker, D. (2000) "Frankfurt's Attack on the Principle of Alternative Possibilities: A Further Look," *Philosophical Perspectives* 14: 181–201.

Widerker, D. (2003) "Blameworthiness, and Frankfurt's Argument Against the Principle of Alternative Possibilities," in D. Widerker and M. McKenna (eds), *Moral Responsibility and Alternative Possibilities*. Aldershot: Ashgate, pp. 53–74.

Wyma, K. (1997) "Moral Responsibility and Leeway for Action," *American Philosophical Quarterly* 34: 57–70.

Zagzebski, L. (2000) "Does Libertarian Freedom Require Alternate Possibilities?" *Philosophical Perspectives* 14: 231–48.

Further Reading

Widerker, D. and McKenna, M. (2003) *Moral Responsibility and Alternative Possibilities: Essays on the Importance of Alternative Possibilities*: Aldershot: Ashgate. (The most comprehensive collection of papers to date on Frankfurt-style cases.)

Related Topics

Semicompatibilism
Leeway vs. Sourcehood Conceptions of Free Will