

Unavoidable Actions

Justin A. Capes

Department of Humanities, Flagler College

Forthcoming in *Philosophical Explorations*; please cite published version

Abstract: It's often assumed, especially in discussions of free will and moral responsibility, that unavoidable actions are possible. In recent years, however, several philosophers have questioned that assumption. Their views are considered here, and the possibility of unavoidable actions is defended and then applied to issues in action theory and in the literature on moral responsibility.

Keywords: Ability; Action; Agency; Frankfurt cases; Two-way Powers; Unavoidability

It's often assumed in action theory and in ethics, and in discussions of free will and moral responsibility in particular, that unavoidable actions are possible. In recent years, however, several philosophers have expressed skepticism about whether that assumption is defensible.¹ I'll argue here that it is. More precisely, I'll argue that there is an important sense in which actions can be unavoidable, and I'll apply that conclusion and my argument for it to two further issues, one concerning the nature of action, the other concerning the conditions for moral responsibility.

1. Unavoidability

I begin by explaining the sense in which I think it's possible for actions to be unavoidable. As I understand it, to say that something is *unavoidable* for an agent is to say that the agent lacks an

ability to avoid that thing. So, for example, to say that a particular action like ‘shooting Smith’ or ‘walking the dog’ is unavoidable for an agent is to say that the agent lacks an ability to avoid that action. There are, however, different kinds of abilities, or so it’s commonly thought, and so there are different senses in which an action can be said to be unavoidable for an agent, depending on the kind of ability the agent is said to lack.² For an action to be unavoidable in the sense I have in mind is for its agent to lack what I’ll call an *in-the-moment* ability to avoid the action. So, unless I indicate otherwise, when I mention agents’ abilities in what follows, I mean their in-the-moment abilities. Likewise, when I talk about what agents can or can’t do.

In-the-moment abilities are, as the label suggests, abilities to do things in the moment, right then and there, in the specific situation in which one finds oneself, or in some specific future situation. (For this reason, they are sometimes referred to as *specific* abilities.) They differ from what are sometimes known as *general* abilities, which an agent can retain even in situations in which the agent isn’t in a position to exercise them.³ Very roughly, having a general ability to perform a certain action *A* is a matter of having the knowledge and skill necessary for *A*-ing, whereas having an in-the-moment ability to *A* requires, in addition, being in a position to put that knowledge and skill into action on some specific occasion. As John Martin Fischer puts it, an in-the-moment ability is “the sort of ability that corresponds to J.L. Austin’s ‘all-in’ sense of ‘can’ or ‘can in the particular circumstances’” (2002, 304).⁴

To illustrate the distinction, consider a pianist trapped in an empty room. Despite her predicament, the pianist retains a general ability to play the piano, insofar as she retains the requisite musical knowledge and skill. Being locked in an empty room doesn’t deprive her of those things. However, because of her predicament, she can’t put that knowledge and skill into action right this moment and so lacks an in-the-moment ability to play the piano now.

Examples like this might suggest that whether an agent has an in-the-moment ability to A depends solely on whether the agent has a general ability to A and whether circumstances external to the agent are conducive to the exercise of that general ability. But the suggested view is incorrect. Last night at 2:00 a.m., Jill, an accomplished concert pianist, had a general ability to play the piano and circumstances external to Jill were conducive to her exercising that ability. But because Jill was fast asleep at 2:00 a.m., she wasn't in a position just then to play the piano and so lacked an in-the-moment ability at that time to play the piano then. As this example illustrates, whether an agent has an in-the-moment ability to A doesn't depend solely on whether the agent has a general ability to A and whether circumstances external to the agent are conducive to the exercise of that ability. It also depends on intrinsic features of the agent other than those that ground the agent's general ability to A .⁵ Bear that in mind as we proceed.

2. An Unavoidable Action

I'll now argue that it's possible for actions to be unavoidable in the sense of 'unavoidable' just adumbrated. In other words, I'll argue that it's possible for an agent who performs an action A at time t to lack an in-the-moment ability to avoid A -ing at t .⁶ I'll start by arguing that it's possible for an agent to find herself in a situation in which she lacks an in-the-moment ability to *intentionally* avoid A -ing in that situation, but later I'll briefly explain how my argument for that claim can be extended to show that it's possible for an agent to find herself in a situation in which she lacks an in-the-moment ability to avoid A -ing in that situation, period.

My argument for the possibility of unavoidable actions features situations of the following sort: an ordinary, mentally healthy agent wants badly to A , intends to A straightaway,

and has no countervailing motivation. In saying that the agent “has no countervailing motivation” I mean that the agent has zero motivation that inclines her toward not *A*-ing.

Here’s a specific case of the sort I have in mind.⁷ With no regard for her own safety, a mother rushes into a burning house in an effort to rescue her two children who are trapped inside. Upon learning that her children were trapped inside, the mother wanted very badly to rush in and save them, she intended to do so straightaway, and she had no countervailing motivation whatsoever. Indeed, the thought of not rushing into the house never even occurred to her, so overwhelmed was she by the thought of her beloved children perishing in the blaze.

An agent in a situation like this might intentionally avoid *A*-ing, but for that to happen the agent would have to undergo a non-trivial change in motivation. Either she would have to acquire some motivation not to *A*, on the basis of which she might then change her mind about *A*-ing, or she would have to get rid of or otherwise lose at least some of her existing motivation to *A* and with it the intention she has to *A* straightaway. Absent any such change in motivation, though, it’s hard to see how the agent in a case like this might intentionally avoid *A*-ing.⁸

That’s because intentional omissions, like intentional actions, are explicable. More specifically, if an agent intentionally refrains from doing something she was strongly motivated to do, intended to do, and had no motivation not to do, there should be an explanation of why that is. Consequently, if the mother in the case at hand had deliberately refrained from rushing into the burning house despite her intention to rush in and despite initially not having any motivation to refrain from rushing in, there should be some explanation of why she did so.⁹

So, what might explain the mother’s change of course in the counterfactual scenario in which she intentionally refrains from rushing in? Any answer to this question looks like it’s going to refer to a change in the mother’s motivational states—for example, the waning of the

initial desire to save her children or the acquisition of a new reason not to rush in the recognition of which motivates her to refrain from rushing in. But without *some* such motivational change, the mother's *intentionally* not rushing in would seem to be totally inexplicable, given that she intended to rush in and had no motivation not to do so. So, unless one wants to say that the mother intentionally not rushing in could be totally inexplicable, it looks as if the mother would have had to undergo a change in motivation in order to intentionally not rush into the house.

To this it could be objected that the mother's intentionally not rushing in *could* be totally inexplicable, if "explicable" is understood in a strong way such that "even though she really wanted to, she just didn't rush in" doesn't count as a satisfactory explanation. After all, we say things like this all the time, don't we? For example, when a coworker asks you what you did this weekend, you might say, "Well, I really wanted to go to the movies, but I didn't." If the coworker follows up by asking why you didn't go, you might say, "I don't know; I just didn't."¹⁰

Even in cases like this, however, there is presumably some explanation of why you didn't perform the relevant action, some motive or desire that, if pressed, we could identify that influenced your behavior. When we say things like "I really wanted to go but didn't," we seem to be signaling not that our omission was totally inexplicable but rather that there was no *special* or *noteworthy* explanation. It's not that you refrained from going to the movies because you got sick or had to finish up that project for work or had to attend a friend's birthday party. The explanation, rather, is more pedestrian—something like, you just wanted to stay home.

I've argued that the mother who rushes into the burning house to rescue her children would have had to undergo a change in motivation in order to intentionally not rush into the house. What are we to say, then, if it was impossible, given the circumstances in which the mother found herself, for her to undergo the necessary motivational change? Suppose, for

example, that there was nothing she could have done to motivate herself not to rush into the house or to rid herself of her intention to rush in (e.g., had she paused to remind herself of how dangerous running into burning buildings is, that wouldn't have dissuaded her). Suppose, too, that there was no way for the requisite motivational change to occur independently of the mother's agency (e.g., there was no one else around who might have said something that would have made her think twice about rushing into the burning house, nor was it psychologically possible at the time for the women to undergo a spontaneous motivational change). In short, suppose that the mother was "stuck" with the motivational structure she found herself with at the time. (I'll defend the legitimacy of this supposition momentarily.) Then I think we should say that the mother was unable (lacked an in-the-moment ability) to intentionally avoid rushing in.

To see why I say this, compare the mother in the case we have been discussing to Jonah, a college student who has an important assignment due tomorrow in one of his classes. To complete the assignment, Jonah needs a certain book from the library. Does Jonah now have an in-the-moment ability to complete the assignment on time? That depends, I should think, on whether it's now possible for him to procure the needed book. Suppose it isn't (e.g., because the book is checked out by someone else). Then it seems accurate to say that it isn't now up to Jonah whether he completes the project on time, that he isn't now in a position to complete it on time, and, accordingly, that he now lacks an in-the-moment ability to do so.

A similar line of reasoning applies to the mother who rushes into the burning house. Did she have an in-the-moment ability to intentionally avoid rushing into the house? That depends, I should think, on whether it was possible for her to undergo the necessary motivational change. If it wasn't, then it seems accurate to say that it wasn't up to her at the time whether she rushed into

the house, that she wasn't in a position at the time to avoid rushing in, and, accordingly, that she lacked an in-the-moment ability to avoid doing so.

The operative principle in both cases is something like this: if an agent needs x to do y and circumstances preclude the agent from acquiring x , then the agent isn't currently in a position to do y and so lacks an in-the-moment ability to do y . Jonah needs the book in order to complete his assignment on time, but, given the circumstances, it's not possible for him to obtain the book, which is why he can't complete the assignment on time. Similarly, I've argued that because the mother was so strongly motivated to rush into the house to save her children, intended to do so straightaway, and had no countervailing motivation, she needed a change in motivation in order to intentionally refrain from rushing into the burning house. But we are assuming that circumstances precluded her from undergoing the requisite motivational change. It follows from all this that she wasn't in a position at the time to intentionally avoid rushing into the house and so lacked an in-the-moment ability at that time to intentionally avoid rushing in.

Readers will note that my argument for this conclusion relies heavily on the supposition that, as I put it earlier, the mother was "stuck" with the motivational structure she found herself with at the time. Indeed, the assumption is essential to my argument, for without it, the argument would be guilty of a fallacy discussed by John Martin Fischer and Mark Ravizza (1992).¹¹ The fallacy involves moving from the premise that an agent S lacks something x that is necessary for S to do y to the conclusion that S is unable to do y . But, as Fischer and Ravizza point out, the conclusion doesn't follow from the premise, as it may be possible for S to obtain x and to thereby put himself in a position to do y . If it is, then it would seem that the agent is now able to do y (at a later time) even though he now lacks something necessary for doing y (at that later time).

So, for example, from the fact that Jonah doesn't now have the materials he needs to complete his assignment on time, we can't validly infer the conclusion that Jonah can't complete the assignment on time. To be justified in inferring that conclusion, we would also need the premise that Jonah can't obtain the requisite materials in time. Similarly, from the fact that the mother doesn't currently have the motivational structure necessary for her to intentionally refrain from rushing into the house, we aren't warranted in concluding that she is unable (in the moment) to intentionally refrain from rushing in. To be justified in that conclusion, we would also need to suppose that it's impossible, given the circumstances, for her to acquire the requisite motivational structure. We should thus examine that supposition more closely before proceeding.

Against the supposition, it could perhaps be argued that an agent's motivational structure at any given time is always malleable, so that agents will always be in a position to acquire, whether actively or passively, sufficient motivation to behave differently. If so, then it would be illegitimate for me to simply stipulate, as I have done thus far, that the mother couldn't have undergone the requisite motivational change. But the claim that our motivation must always be pliable in this way is highly implausible. Surely, it's at least *possible* for agents to find themselves in situations in which their motivational structures are unalterable. Indeed, I would go further and say that we sometimes do find ourselves in situations of this sort, situations, that is, in which we have beliefs, desires, values, preferences, and intentions that we can't immediately rid ourselves of while lacking others that we can't immediately acquire, and that this, in turn, results in our being "stuck," for a time, with the relevant motivational structures.

I take no stand here on how frequent such situations are. Fischer and Ravizza (1992, 434) have argued that they are infrequent, whereas Peter van Inwagen (1994, 99-106) contends that they are fairly common. My suspicion is that the truth lies somewhere in between, but it doesn't

much matter for present purposes whether I'm right about that. My claim at present is simply that there can be (and likely are) situations in which we are stuck, at least for a time, with our existing motivational structures, a claim that strikes me as highly plausible.

In support of the claim, consider another example. Someone takes me to a window on the tenth story of a building and says, "I'll give you five bucks if you jump out of this window onto the empty parking lot below within the next minute." There is no way that I would do it. Indeed, I would go further. Not only would I not agree to jump out of the window for a measly five dollars, I'm confident that I couldn't bring myself to do so. That's not to say that there are no circumstances whatsoever in which I could be induced to jump out of a tenth story window. Perhaps if the building were on fire, or my family were being threatened, or if it was the only way to prevent a nuclear holocaust, or something like that. But for five bucks? In the next sixty seconds? There's just no way. Given my current motivational structure, I couldn't bring myself to jump out of the window for that paltry sum in that short time frame.

To intentionally jump out of the window within the next minute, it seems that I would need some motivation to do so. I currently have none. Could I, within the relevant time frame, actively motivate myself to jump? I doubt it. It's not implausible to suppose that no matter what I told myself or how much I tried to psych myself up to jump out of the window, none of those actions would result in my acquiring sufficient motivation to throw myself out of a tenth story window just to earn five dollars. I plausibly have certain beliefs, desires, values, preferences, and intentions that I couldn't rid myself of in sixty seconds and that prevent me from throwing myself out of a window in the next minute just to earn five bucks. If so, then there is nothing I could do (within the relevant time frame) to acquire the necessary motivation to jump.

There are, of course, variations of this scenario in which my motivational structure might be altered in such a way that I could be induced to jump out of the window for such a tiny amount of money. Perhaps I could acquire the necessary motivation by taking a pill that would make me go insane. But we can safely assume that I don't have access to any such pill and, more generally, that, circumstances being what they are, there is nothing that could happen to me in the relevant time frame that would produce in me the necessary motivation to jump. Given these suppositions, it looks like I'm stuck with my existing motivational structure, a structure that precludes me from intentionally flinging myself out of the window onto the pavement.

This, of course, is a fictional case, but no doubt there are actual cases like it in which an agent not only lacks the motivation necessary to perform a certain action, but in which it's also impossible, given the circumstances, for the agent to acquire the requisite motivation.¹² Surely, then, there could be parallel situations, like the story about the mother and the burning house, in which an agent lacks the motivation necessary to intentionally *avoid* performing an action she fully intends to perform, and in which it's also impossible, circumstances being what they are, for the agent to acquire the requisite motivation to intentionally avoid performing that action. If so, then the supposition that the mother was stuck with the motivational structure with which she found herself seems unobjectionable. But, as I argued above, if the mother was stuck with that motivational structure, then she was unable to intentionally avoid rushing into the house.

I want to pause briefly to summarize the argument up to this point. I began by arguing that there are possible cases in which an agent intentionally performs an action *A* and in which, given the agent's intentions and motivational structure at the time, the agent would need a significant change in her motivational structure to intentionally refrain from *A*-ing. I've also argued that in such cases it may be impossible, circumstances being what they are, for the agent

to acquire the needed motivational change. But, as we have seen, if an agent needs x to do y and circumstances preclude the agent from acquiring x , then the agent isn't currently in a position to do y and so lacks an in-the-moment ability to do y . It follows from all this that there are possible cases in which an agent who intentionally performs a specific action A (such as the mother who rushes into the burning house in an effort to save her children) isn't in a position to intentionally avoid A -ing and so lacks an in-the-moment ability to intentionally avoid A -ing.

Suppose I'm right that there is a possible version of the story about the mother trying to save her children in which the mother couldn't have *intentionally* avoided rushing into the burning house. That leaves open the possibility, however remote, that the mother could have *non-intentionally* avoided rushing in (e.g., by tripping and suffering an incapacitating injury while running toward the house). Rushing into the house therefore wasn't completely unavoidable for her. However, we can easily augment the case in ways that would render the mother's action completely unavoidable. For example, we might suppose that there was a guardian angel on the scene whose sole responsibility was to prevent the mother from non-intentionally not rushing in (e.g., by making sure there were no toys in her path on which she might trip). Making this addition to the story yields a possible case in which an action was completely unavoidable for its agent in the sense that the agent lacked, at or immediately prior to the time of action, an in-the-moment ability to avoid performing that action in that situation.

3. Alvarez's Analysis

It's instructive to compare my analysis of the case we have been discussing to that of Maria Alvarez (2013). Alvarez argues that the mother in the example was in fact able to avoid rushing into the house and, consequently, that the case isn't an instance of unavoidable action. As we'll

see, however, while Alvarez's argument may support the conclusion that the mother had a *general* ability to avoid rushing in, it doesn't support the conclusion that she had an *in-the-moment* ability to avoid doing so and thus doesn't challenge my analysis of the case.

Alvarez claims, plausibly, that the mother would have refrained from rushing into the house had she been "told that by running into the house she'd make it more likely that her children will die... This," Alvarez contends, "shows that [the mother] had the ability and opportunity to refrain in the actual case and that it was up to her how she acted." Alvarez goes on to say that what the mother lacked but would have had had she been informed that rushing into the house would further endanger her children "is the motivation to refrain from rushing into the house." And Alvarez contends that lacking the motivation to refrain from rushing in doesn't render the mother unable to avoid rushing in. To insist that it does, she says, would be to treat "motives (reasons, desires, inclinations) to act otherwise as conditions for being able to act otherwise, which seems wrong," as it conflates the "conditions for the exercise of an ability with the conditions for the presence of the ability" (2013, 117).

There are several aspects of these remarks that I want to discuss, starting with Alvarez's inference from the fact that the mother would have refrained from rushing into the house had she been informed that doing so would "make it more likely that her children will die" to the conclusion that "it was up to [the mother] how she acted." I grant the premise, but the conclusion doesn't follow. To see this, consider again the case of Jonah, the college student who has an important assignment due tomorrow. Let's suppose that Jonah would finish his assignment on time if he were given that one book from library that he needs. Does it follow that it's up to Jonah whether he finishes the assignment on time? It does not, for suppose he has no way of getting the needed book and that no one is willing to give it to him. In that case, it evidently

follows (given the assumption that Jonah must have the book to finish the assignment) that it's not now up to Jonah whether he completes the assignment on time and that he lacks an in-the-moment ability to complete the assignment on time.

Similar remarks apply to the case of the mother. Let's suppose, with Alvarez, that the mother would have refrained from rushing into the house had she been informed that rushing in would further endanger the lives of her children. By itself that doesn't establish that it was up to the mother how she acted or that the mother had an *in-the-moment* ability to avoid rushing into the house, just as the fact that Jonah would complete his assignment on time if he had the necessary resources doesn't establish that he has an in-the-moment ability to complete the assignment on time. For suppose that the only way the mother could have brought herself to forbear rushing into the house is if she had learned that doing so would increase the risk to the lives of her children. Suppose, too, that there was no one around to inform her of the increased danger to her children that her rushing into the house would occasion, and that she was incapable of recognizing the increased danger herself. In that case, it seems correct to say, for reasons given in section 2, that it wasn't up to the mother whether she intentionally rushed in, that the mother wasn't in a position to intentionally avoid rushing in, and, accordingly, that she lacked an in-the-moment ability to intentionally avoid rushing in. And it seems correct to say all that even if it's true that the mother would have intentionally refrained from rushing in had she been informed that her running into the house would increase the risk to her children.

The next thing to note about Alvarez's analysis of the case is that the ability to refrain that she attributes to the mother isn't an in-the-moment ability. It's a general ability, and the opportunity to refrain a matter of the agent's external circumstances being conducive to the exercise of the attributed general ability (2013, 108). Thus construed, Alvarez's claim that the

mother “had the ability and opportunity to refrain” from rushing into the house is no doubt correct. The mother presumably retained a general ability to avoid rushing into the house at that time, and if we stipulate that an agent’s opportunities to act are determined solely by her external circumstances, the mother had the opportunity to refrain as well, since her external circumstances, considered in and of themselves, independently of her psychology, didn’t prevent her from exercising her general ability to refrain. There may, then, be a sense in which the mother could have avoided rushing into the house. She had the general ability to do so, and her external circumstances allowed for the exercise of that ability. However, that’s compatible with there being another sense in which she couldn’t have avoided rushing into the house, a sense captured by the claim that she lacked an in-the-moment ability to avoid rushing in.

To illustrate, consider Jill again, the concert pianist discussed in section 1. You’ll recall that, although Jill was sound asleep at 2:00 a.m., she retained at that time a general ability to play the piano, and her external circumstances afforded her the opportunity to exercise that ability. But it doesn’t follow that Jill had an in-the-moment ability at 2:00 a.m. to play the piano then. In fact, she lacked that ability. Because Jill was asleep at the time, she wasn’t in a position just then to play the piano then and so lacked an in-the-moment ability to do so.

Similar remarks apply to the mother who rushed into the burning house. She may have retained a general ability at the time to intentionally avoid rushing into the house, and her external circumstances may have afforded her the opportunity to exercise that general ability. However, it doesn’t follow that she had an in-the-moment ability to avoid rushing in, and, as I have argued, it’s plausible, given that she was “stuck” with motivational structure she had at the time, that she didn’t have an in-the-moment ability to avoid rushing into the house. Perhaps, then, what we should say about the mother is this: while she had a general ability to avoid

rushing into the house, and while her external circumstances at the time afforded her the opportunity to exercise that ability, she just couldn't bring herself to exercise it in this instance.

Consider, finally, Alvarez's remarks about the relationship between motivation and ability. Alvarez grants that the mother lacked "the motivation to refrain from rushing into the house," but argues that lacking the motivation to refrain doesn't render the mother unable to avoid rushing in. To insist that it does, Alvarez thinks, would be to treat the motivation to refrain as a condition for being able to refrain, which she thinks would be a mistake, conflating the "conditions for the exercise of an ability with the conditions for the presence of the ability" (2013, 117).¹³ This is certainly true when it's a general ability that's at issue, and one might think that similar things could plausibly be said about in-the-moment abilities as well.

It may indeed be a mistake to treat the motivation to refrain as a condition for being able (i.e., having an in-the-moment ability) to refrain. But if so, it's not a mistake I've made in arguing for the possibility of unavoidable actions. Note that it's not just the fact that the mother lacked the motivation to refrain from rushing into the house that leads me to conclude that she couldn't have intentionally avoided rushing in (just as it's not simply the fact that Jonah lacks the needed book that leads me to conclude that he can't now complete his assignment on time). It's also the fact that it was impossible, in the circumstances, for the mother to acquire the requisite motivation. It's one thing to say that an agent might retain an in-the-moment ability to intentionally avoid an action she fully intends to perform despite not being at all motivated to avoid performing that action. It's something else entirely to say that the agent might retain that ability even though it's impossible for her to acquire the motivation that would be necessary for her to intentionally avoid the action. It would be much like saying that Jonah is now in a position

(and so has an in-the-moment ability) to complete his assignment on time even though the resources he needs to do so are currently unobtainable, and it would be just as implausible.

4. Detailed Unavoidability

Suppose one were to concede that, when described in certain very general ways, actions can be unavoidable for their agents. Even so, it would still be possible to maintain that *something* about an agent's behavior must be avoidable for the agent if the behavior in question is to qualify as genuine action of that agent. Helen Steward (2012), for example, acknowledges that there might come a point at which a compulsive handwasher is "strained to the breaking point," so that the person simply has to wash her hands "and could not leave it another moment." However, as Steward goes on to point out, in a typical case of that sort, the agent still "retains a range of (admittedly minimal) freedoms to act in one way rather than another. She can wash with soap or not, with hot water or cold, for one minute or a bit longer, just the hands or up to the elbows...moving one's hands in this way or that way, etc." (2012, 183-184). These things, it seems, remain up to the agent even if, owing to her compulsion, it isn't up to her at the time whether she washes her hands on this occasion. And according to Steward, the handwashing qualifies as an action only if the person retains at least some of the minimal freedoms just mentioned. On Steward's view, then, while unavoidable actions might be possible (at least when the actions are described in very basic ways like 'washing one's hands'), what's not possible are instances of agency every aspect of which is completely unavoidable for the agent.

One nice thing about Steward's position is that it accommodates seemingly competing intuitions about unavoidable actions. By insisting that at least some aspects of an agent's behavior must be up to the agent if the behavior is to qualify as an action, the view gives

expression to thoughts about choice and control that often motivate skepticism about the possibility of unavoidable actions. However, the view also allows us to acknowledge that there may well be cases like the mother who rushes into the burning house or Steward's compulsive handwasher in which an agent is unable to avoid acting in a certain general way. This advantage notwithstanding, doubts can be raised about the cogency of Steward's view.

What if Steward's compulsive handwasher has a painstakingly detailed routine, so that she is especially fastidious about the order of the various activities that constitute handwashing and the speed at which they are carried out? Mightn't even these details of her behavior be unavoidable for her, for much the same reason as is her more general handwashing behavior? I don't see why not. But provided the behavior in question is a non-deviant result of the woman's desire to wash her hands and her intention to wash them at that time and in that way, denying that the handwashing is among the woman's actions would be controversial at best.

Perhaps, though, you find that the sort of case we are now attempting to envision, in which even the precise details of the agent's behavior are unavoidable, strains the imagination. Steward can empathize. She finds it "very hard indeed properly to imagine an agent subject to such constraints" (2012, 185). I suspect, though, that the difficulty stems largely from the fact that we are concentrating on cases involving behaviors like handwashing that consist of multiple steps which take several moments to carry out and that are performed in situations allowing for deliberation and forethought. When we focus on cases like that, the claim that even the precise details of an agent's behavior might be unavoidable can easily seem like a stretch. But the difficulty becomes significantly less acute when we focus on cases that lacks these features.

Consider, for example, what we might think of as simple reflex actions—those done more less automatically in response to external stimuli. The typical driver who must immediately

slam on his brakes to avoid rear-ending the car in front of him will do so automatically, without even thinking about it, as we say. And, indeed, what is there to think about? It's plausible that, in a case like this, the driver might be unable (for the sorts of reasons articulated in section 2) to avoid slamming on his brakes. It's also plausible that agents in cases like this have little choice about the precise details of how and when they act. There simply aren't that many details about which the agent might have a choice. The relative simplicity of the action, the strength of the agent's motivation to perform it, and the urgency with which the agent must act seemingly work in concert to preclude the sort of detailed ability to refrain that Steward insists is necessary for action. Yet reflex behaviors like slamming on one's brakes are intentional actions.

I've encountered people who deny this last claim, who insist that reflex behaviors are more akin to an involuntary twitch than to intentional actions. But note that, unlike an involuntary twitch, reflex behaviors like slamming on one's breaks are often intentional and done for reasons. The average driver who slams on his breaks in response to something happening in front of him intends to slam on his breaks and slams them on intentionally and for a reason (e.g., to avoid hitting the child that has just darted into the street). The fact that the driver's behavior in such cases is intentional and done for reasons strongly suggests that it's an action of his, something of which he is the agent, and not a mere bodily movement like an involuntary twitch which is neither intentional nor done for reasons. So, if such behaviors can indeed be unavoidable, as I've argued they can be, they are plausible candidates for unavoidable actions.

5. Action and Two-way Powers

So far in this paper, I've clarified the claim that unavoidable actions are possible and defended an argument for it. In this section and the next, I apply that claim and my argument for it to two further issues, the first of which concerns the nature of action and agency.

Some philosophers think that acting essentially involves exercising a two-way power—a power “to act or to refrain from acting,” as Steward (2012, 155) puts it—so that what someone does counts as an action only if the person could have refrained from doing it. For example, in the course of discussing whether one person could cause another person to perform an action in a way that would render the action unavoidable, Alvarez claims that “the concept of what it is for someone to act, i.e. for something to be an action of which one is the agent, makes the idea that it is possible to cause the relevant actions so that they are unavoidable problematic.” Why? Because “It seems that for what someone does to be *his action*, for him to be its *agent*, the person must have a certain degree of control over it: at least, he should be capable of refraining from doing what he does” (2009, 75-76). Hanna Pickard makes similar remarks in the course of criticizing a fairly common conception of psychopathology according to which people with various “disorders of agency” (e.g., addicts, kleptomaniacs, obsessives, etc.) lack the ability to do otherwise. Pickard argues, convincingly, that this conception “is false,” and offers “an empirically and clinically informed understanding of disorders of agency which preserves the ability to do otherwise” (2015, 135).¹⁴ The fact that people who suffer from disorders of agency have the ability to do otherwise fits with what Pickard describes as “our pre-theoretical concept of action,” which “connects [action] to choice and control: action involves the possibility of doing otherwise, at least in so far as one could refrain from performing *that very action*” (155).

What are we to make of such claims in light of the fact that unavoidable actions are possible? We might simply reject them and move on, but that, I think, would be precipitous. As I

noted about the case of the mother who rushes into the burning house, there may well be a sense in which the mother in that case could have avoided rushing into the house (a sense captured by Alvarez's claim that the mother had a general ability to refrain from rushing into the house the exercise of which wasn't precluded by her external circumstances), even though there is another sense in which the mother couldn't have avoided rushing in (a sense captured by the claim that she lacked an in-the-moment ability to avoid rushing into the house). Similarly, it could be, as Alvarez says, that "for what someone does to be *his action*, for him to be its *agent*, the person...should be capable of refraining from doing what he does," provided the capability in question is a general ability. That's consistent with the claim that "for what someone does to be *his action*, for him to be its *agent*, the person" needn't have an in-the-moment ability to refrain. Pickard may also be right that "action involves the possibility of doing otherwise, at least in so far as one could refrain from performing *that very action*," as long as the 'could' here expresses a general ability to refrain from the relevant action and not an in-the-moment ability to refrain.

Initial appearances to the contrary, then, the fact that unavoidable actions are possible doesn't settle the question of whether agency essentially involves the exercise of a two-way power. It does, however, place a constraint on how an affirmative answer to that question should be understood. If agency does indeed involve the exercise of a two-way power, as Steward, Alvarez, and Pickard suggest, the power isn't of a kind that involves an in-the-moment ability on the part of the agent to refrain from behaving as she does.

6. Frankfurt Cases

The preceding discussion of unavoidable actions also suggests some strategies for defending a well-known argument of Harry Frankfurt's against some recent challenges to it. Frankfurt's

argument targets what he calls “the principle of alternate possibilities” (PAP), according to which “a person is morally responsible for what he has done only if he could have done otherwise” (1969, 829). There are various ways of interpreting the phrase ‘could have done otherwise’ in PAP, resulting in various interpretations of the principle. But as I and most others interpret it, the ‘could’ here is the ‘could’ of in-the-moment ability and ‘doing otherwise’ encompasses both performing some other action instead and simply not performing the action one did perform.¹⁵ Thus construed, PAP can be restated as follows: a person is morally responsible for what he did only if he had an in-the-moment ability to avoid doing it.

In an effort to falsify this principle, Frankfurt tells the following story:

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 does become clear that Jones is going to decide to do something else, Black takes effective steps to ensure that Jones decides to do, and that he does do, what he wants him to do. Whatever Jones’s initial preferences and inclinations, then, Black will have his way...Now suppose that Black never has to show his hand because Jones, for reasons of his own, decides to perform and does perform the very action Black wants him to perform.” (1969: 835-836)

To fill in the example, imagine that what Black wants Jones to do is to kill some other guy, Smith. Frankfurt contends that Jones couldn’t have avoided killing Smith in this case, but that,

because Jones killed Smith for reasons of his own, he can be morally responsible for his action, nonetheless. If Frankfurt is right about that, then we have a counterexample to PAP.

I want to focus on the first premise of Frankfurt's argument, the claim that Jones couldn't have avoided killing Smith in this case. Why think that claim is true? Note that it won't do to simply point out that Jones would have killed Smith as a result of Black's efforts had he not killed Smith for reasons of his own. As Alvarez correctly points out, "unless a premise is added to the effect that Black's intervention is irresistible, this reasoning involves a modal fallacy, for, from a premise that something will happen, it is not legitimate to infer that nothing else could have happened" (2009, 67). It appears, then, that Frankfurt must also stipulate that Black's efforts to bring it about that Jones kills Smith would have been irresistible by Jones.

According to Alvarez, though, Frankfurt isn't entitled to any such stipulation. She says, "it is not legitimate for a Frankfurt-style case simply to *stipulate* that, in the counterfactual case, the agent would be caused to perform an action that he cannot avoid performing... Rather, any example needs to tell a compelling story that makes the suggestion plausible without begging the issues at hand" (2009, 67). Frankfurt makes some suggestions about how to tell such a story, but Alvarez finds his suggestions problematic, as do several other recent critics of his argument.

One of Frankfurt's suggestions is that Black could pronounce a threat so terrible that it would compel Jones to kill Smith (1969, 835). In response to this suggestion, Alvarez contends that threats work "by making non-compliance... highly unpalatable to the agent—not by eliminating its possibility" (2009, 72). No doubt that's true in many cases, but mightn't some threats be so severe and make non-compliance so extremely unpalatable to the agent that the agent simply can't help complying with the threat? I don't see why not.

Imagine a bank teller being robbed at gunpoint. The gunman threatens the teller's life if she doesn't hand over the money, and the teller knows that the gunman means business. This same band of thieves has been robbing banks all over the city, and just last week an intrepid teller at another institution across town refused to comply with the gunman's demand and was shot down in cold blood for her temerity. Knowing this, and not being especially bold by nature, the teller currently under the gun is overwhelmed with fear and sees no reason not to comply with the gunman's demand. Calling the gunman's "bluff" is, for her, in these circumstances, simply unthinkable. Accordingly, she intentionally hands over the cash post-haste.

Was it within the teller's power in this case to avoid handing over the cash? Perhaps, but perhaps not; the devil, as we know, is in the details. However, I've no doubt that we could spell out the pertinent details of the story so that it mirrors the cases of unavoidable action discussed above (e.g., by stipulating that the teller's motivational situation is much like that of the mother considered earlier who rushes into a burning house to save her children who are trapped inside). But if we can do that with this case, we can surely do it with Frankfurt's case as well.

It could be objected that if the teller was so overwhelmed with fear that she truly couldn't help complying with the threat, then the resulting event (handing over the money) wouldn't really count as her action. Although it would be something that the agent did, it wouldn't be something attributable to the agent at the personal level and thus wouldn't really be *her* action (cf. Alvarez 2013: 113-114; see also Alvarez 2009).¹⁶ To insist otherwise, it might be, said begs the question against those who insist that actions must be avoidable for their agents.¹⁷

I agree that simply to assert that unavoidable behaviors are genuine actions would beg the question. However, I think we have reason to think that if handing over the cash really was unavoidable for the teller given her motivational structure at the time, that it was still an action of

hers. Even if the teller was so overwhelmed by her fear that she automatically complied with the threat and couldn't have done (i.e., didn't have an in-the-moment ability to do) otherwise, complying is still plausibly something she did intentionally and for a reason (e.g., to avoid being shot by the robbers). But, again, the fact that the teller's behavior was intentional and done for reasons strongly suggests that it was an action of hers, something of which she is the agent, and not a mere bodily movement like an involuntary twitch or moving while asleep, behaviors that are neither intentional nor done for reasons. So, if such behaviors can indeed be unavoidable, as I've argued they can be, they are plausible candidates for unavoidable actions.

Another one of Frankfurt's suggestions, the one that has captured people's philosophical imagination and which is now the standard answer to how Black could ensure that Jones kills Smith, involves a bit of science fiction. We are to imagine that Black is a nefarious neuroscientist who has secretly implanted a device in Jones's brain that allows him to read Jones's thoughts and, if need be, control Jones's behavior. Had it become clear to Black that Jones wasn't going to kill Smith on his own, Black would have used the device to "manipulate the minute processes of Jones's brain and nervous system...so that causal forces running in and out of his synapses and along the poor man's nerves determine that he" kills Smith (1969, 835-836).

In response to this story, several philosophers have expressed skepticism about whether it's possible for Black, using his device, to cause Jones to perform an action at all, much less the same action that he performs in the actual sequence of events (see, e.g., Alvarez 2009, Larvor 2010, and Steward 2008, 2009). If their doubts are well-founded, if what Black causes to happen in the counterfactual sequence of events in which he intervenes and manipulates Jones's brain isn't an action of Jones's, then it would seem that Jones could have avoided the action he performed after all, in which case we obviously don't have a counterexample to PAP.

It seems to me, however, that such skepticism about the causal powers of Black and his device is unwarranted. If, as some believe, the sorts of mental states that move us to action (beliefs, desires, intentions, etc.) just are or supervene on or are otherwise essentially tied to specific neural activity in such a way that the occurrence of the relevant neural states ensures the occurrence of those mental states, then I see no barrier to Black using his device to cause Jones's action. And, even if the relevant mental states aren't *essentially* tied to specific neural activity, it's plausible that there is, or could be, a sufficiently tight causal connection between certain types of neural activity, on the one hand, and motivational states like wanting and intending, on the other, that Black, by causing the relevant types of neural activity, could induce in Jones the sorts of motivational states and events that would, in turn, lead to Jones's action.

There are, moreover, other ways for Black to use his device to ensure that Jones kills Smith, ways that don't require Black to initiate Jones's action. To see this, consider the following variation on Frankfurt's original sci-fi case. Imagine that Jones intends, independently of anything Black has done, to kill Smith and that he executes that intention (along with Smith, of course) without hesitation. However, if Jones had hesitated in any way to carry out his intention to kill Smith, Black, with the help of his device, would have picked up on this and, using the device, would have intervened to create a motivational context in which killing Smith is unavoidable for Jones. Having read the earlier parts of this paper, Black knows that all he needs to do to create a situation like that is to make it impossible for Jones to undergo the motivational change necessary for him to intentionally avoid carrying out his existing intention to kill Smith. This, it seems, is something Black could easily do using his device, for example, by "deactivating" the neural pathways that would be involved in Jones being sufficiently motivated

to avoid killing Smith. In doing so, Black wouldn't be 'inducing' Jones to act. He would, instead, simply be making sure that Jones stayed the course that Jones had already set himself on.¹⁸

Does Jones have an (in-the-moment) ability, in this version of the story, to intentionally avoid killing Smith? Arguably, he doesn't. For him to intentionally avoid killing Smith, he would have to undergo a non-trivial change in motivation. But Black is there to make sure that that doesn't happen. It's thus plausible that Jones isn't in a position to avoid killing Smith and, consequently, that he lacks an in-the-moment ability to avoid killing Smith in that situation.

Note that, in this version of Frankfurt's story, in both the actual sequence of events in which Jones kills Smith on his own and in the counterfactual sequence in which Black intervenes, Jones kills Smith in part because of his intention to do so, an intention formed or otherwise acquired independently of Black's efforts. Note in particular that in the event that Black intervenes, he doesn't force Jones to do anything Jones doesn't already want and intend to do. What Black does, rather, is to ensure a motivational context in which Jones's preexisting intention to kill Smith is inevitably carried out. There is, then, a clear sense in which Jones, not Black, initiates the action, even if Black must intervene to ensure that Jones goes through with it.

So, it seems that both Frankfurt's suggestion about threats and his science fiction story provide the basis for "a compelling story" about how Black could ensure that Jones kills Smith in the event that Jones doesn't kill Smith for reasons of his own. I conclude that whatever problems there may be with Frankfurt's argument against PAP (and I think there are plenty of them), doubts about the possibility of causing unavoidable actions aren't among them.

7. A Further Implication

I've argued that unavoidable actions are possible. More precisely, I've argued that it's possible for an agent to lack an in-the-moment ability to avoid performing a certain action, and I've highlighted some implications of this conclusion and my argument for it for issues in action theory and in the literature on moral responsibility. I'd like to conclude by briefly pointing out one further implication of my argument.

Note that the argument doesn't rely on any far-off metaphysical possibilities. Except for a lone guardian angel introduced at the end of section 2, the sorts of cases on which the argument is based don't involve any nefarious neurosurgeons, diabolical demons, outlandish impairments to a person's agency, or any other such philosophical fictions. (Note, too, that the imagined guardian angel plays only a very minor role in the overall argument.) Nor do the cases involve familiar impairments to a person's agency like addiction, kleptomania, or obsessive-compulsive disorder. The conditions for unavoidable action that I've identified could easily obtain in the actual world and in mentally healthy agents. There is, then, a very real possibility that at least some subset of our actions might be unavoidable. If so, if unavoidable actions are not only possible but perhaps also actual, then debates about the ethical significance of such actions could turn out to be more important and directly applicable than we might initially have supposed.

Notes

¹ See, e.g., Alvarez (2009; 2013), Pickard (2015), and Steward (2008, 2009, and 2012).

² For an overview of the literature on agents' abilities, see Clarke (2015).

³ Different philosophers use different terminology to mark this distinction, but the distinction itself, or one near enough to it, is common. See, e.g., Berofsky (2002, 196), Clarke (2015, 893-

894), Fischer (2002, 304), Mele (2003, 447), van Inwagen (1983, 12-13), and Whittle (2010, 2-3).

⁴ An anonymous referee worries about whether in-the-moment abilities are genuine abilities or whether they are general abilities together with the conditions for the exercise of a general ability. I don't share this worry, but suppose it's correct. In that case, I could rephrase the paper's central thesis as follows: it's possible for an agent who performs an action *A* at *t* and who has a general ability at *t* to refrain from *A*-ing at *t* to be in a situation at *t* in which it's impossible for the conditions for the exercise of that general ability to be realized at *t*. Rephrasing the claim in this way, though more cumbersome, wouldn't alter the substantive points made in what follows.

⁵ It's worth noting at this point that having an in-the-moment ability to *A* may not depend on having a general ability to *A*. Cyr and Swenson (2019) argue that one can have what I'm calling an in-the-moment ability to *A* (and what they call a specific ability to *A*) without having a general ability to *A*.

⁶ My argument for this claim develops some themes from Capes (2012) and is similar in certain respects to an argument of Peter van Inwagen's (1989). An important difference between van Inwagen's argument and mine is that van Inwagen's argument relies on a controversial inference rule, Beta-prime, whereas my argument doesn't.

⁷ The case is discussed by Alvarez (2013, 117) who attributes it to Steward (2012, 183). However, the example appears to have first been introduced by Wolf (1980, 152-153). I discuss Alvarez's treatment of the case in section 2 below.

⁸ For similar points, see Fischer (1994, 52-54), Fischer and Ravizza (1992, 433-434), Wolf (1980, 152-153), and van Inwagen (1994, 99-106).

⁹ The explanation needn't be contrastive. That is, it needn't explain why, in the counterfactual scenario, the mother refrains from rushing into the burning house rather than rushing in.

Depending on one's view of contrastive explanation, it may be that there is no such explanation of the mother's refraining in the counterfactual scenario. The explanation must, however, make intelligible why the mother refrained from rushing in in that case.

¹⁰ Thanks to an anonymous referee for pressing me to address this sort of objection.

¹¹ See also Fischer (1994, 48-62). Fischer and Ravizza attribute this fallacy to van Inwagen (1989).

¹² For discussion of such cases, see Capes (2012, 12-13) and Dennett (1984, 133).

¹³ Note that this mistake is the same one that Fischer and Ravizza (1994, 48-62) attribute to van Inwagen (1989).

¹⁴ It's worth noting that philosophers who, as Pickard (2015, 136) puts it, "turn to psychopathology" when they want "an example of a person who lacks the ability to do otherwise" often don't claim that *actual* agents who suffer from various psychopathologies lack the ability to do otherwise. Their focus, rather, is on hypothetical cases, cases which they take to be conceptually or metaphysically possible. That is certainly what I am up to in Capes 2014.

¹⁵ Fara (2008) is an exception. Fara interprets PAP as a principle about the general ability to do otherwise and shows (conclusively, I think) that Frankfurt's argument has no shot against that version of the principle.

¹⁶ Not everything we "do" counts as a genuine action. Last night you might have woken your spouse with your loud snoring, but waking them, in this case, though something you did, wasn't an action of yours.

¹⁷ Thanks to an anonymous referee for encouraging me to address this objection.

¹⁸ Thanks to an anonymous referee for helpful suggestions about how to formulate this case.

Acknowledgements

Thanks to Randy Clarke, Al Mele, and two anonymous referees for their helpful feedback.

Disclosure Statement

No potential conflict of interest was reported by the author.

Funding

No funding information was reported by the author.

Notes on Contributor

Justin A. Capes is Associate Professor of Philosophy at Flagler College. He works in ethics and the philosophy of action. Most of his research is at the intersection of these two fields.

References

- Alvarez, Maria. 2009. "Actions, Thought-experiments and the 'Principle of Alternate Possibilities'." *Australasian Journal of Philosophy* 87 (1): 61-81.
- Alvarez, Maria. 2013. "Agency and Two-way Powers." *Proceedings of the Aristotelian Society* 113: 101-21.
- Berofsky, Bernard. 2002. "Ifs, Cans, and Free Will: The Issues." In *The Oxford Handbook of Free Will*, edited by Robert Kane, 181-201. New York: Oxford University Press.

-
- Capes, Justin. 2012. "Action, Responsibility, and the Ability to Do Otherwise." *Philosophical Studies* 158 (1): 1-15.
- Clarke, Randolph. 2014. *Omissions: Agency, Metaphysics, and Responsibility*. New York: Oxford University Press.
- Clarke, Randolph. 2015. "Abilities to Act." *Philosophy Compass* 10: 893-904.
- Cyr, Taylor and Philip Swenson. 2019. "Moral Responsibility Without General Ability." *The Philosophical Quarterly* 69 (274): 22-40.
- Dennett, Daniel. 1984. *Elbow Room: The Varieties of Free Will Worth Wanting*. Cambridge Mass: MIT Press.
- Fischer, John Martin. 1994. *The Metaphysics of Free Will: An Essay on Control*. Cambridge, Mass.: Blackwell.
- Fischer, John Martin. 2002. "Frankfurt-type Examples and Semi-compatibilism." In *The Oxford Handbook of Free Will*, edited by Robert Kane, 281-290. New York: Oxford University Press.
- Fischer, John Martin and Mark Ravizza. 1992. "When the Will is Free." *Philosophical Perspectives* 6: 423-451.
- Frankfurt, Harry. 1969. "Alternate Possibilities and Moral Responsibility." *Journal of Philosophy* 66 (23): 829-839.
- Larvor, Brendan. 2010. "Frankfurt Counter-example Defused." *Analysis* 70 (3): 506-508.
- Mele, Alfred. 2003. "Agents Abilities." *Nous* 37 (3): 447-70.
- Pickard, Hanna. 2015. "Psychopathology and the Ability to Do Otherwise." *Philosophy and Phenomenological Research* 90 (1): 135-163.

-
- Steward, Helen. 2008. "Moral Responsibility and the Irrelevance of Physics: Fischer's Semi-compatibilism vs. Anti-fundamentalism." *Journal of Ethics* 12 (2): 129-145.
- Steward, Helen. 2009. "Fairness, Agency and the Flicker of Freedom." *Nous* 43 (1): 64-93.
- Steward, Helen. 2012. *A Metaphysics for Freedom*. New York: Oxford University Press.
- van Inwagen, Peter. 1983. *An Essay on Free Will*. Oxford: Clarendon Press.
- van Inwagen, Peter. 1989. "When is the Will Free?" *Philosophical Perspectives* 3: 399-422.
- van Inwagen, Peter. 1994. "When the Will is Not Free." *Philosophical Studies* 75 (1-2): 95-113.
- Whittle, Ann. 2010. "Dispositional Abilities." *Philosophers Imprint* 10: 1-23.