



Self-Inflicted Frankfurt-Style Cases and Flickers of Freedom

Michael Robinson¹ 

Received: 1 July 2023 / Accepted: 23 October 2023
© The Author(s) 2023

Abstract

According to the most popular versions of the flicker defense, Frankfurt-style cases fail to undermine the Principle of Alternative Possibilities (PAP) because agents in these cases are (directly) morally responsible not for making the decisions they make but for making these decisions *on their own*, which is something they could have avoided doing. Frankfurt defenders have primarily focused on trying to show that the alternative possibility of refraining from making the relevant decisions on their own is not a *robust* alternative, while generally granting that this alternative cannot easily be eliminated from successful cases of this sort. In a recent issue of this journal, Stockdale (2022) attempts to sidestep the debate concerning robustness and develops a novel kind of Frankfurt-style case in which agents are unable to avoid making the relevant decisions on their own. The fundamental problem with Stockdale's argument is that it hinges on an implausible conception of acting on one's own. I help clarify the pertinent sense of what it means to do a thing on one's own in this context and show that these new cases are unable to overcome the targeted versions of the flicker defense of PAP.

Keywords Free will · Moral Responsibility · Alternative Possibilities · Flickers of Freedom · Frankfurt Cases · Acting on One's Own

1 Introduction

The Principle of Alternative Possibilities (PAP) states that a person is morally responsible for what they have done only if they could have done otherwise. Despite its strong intuitive pull and a long history of broad acceptance, many philosophers have

✉ Michael Robinson
michaelrobinson@chapman.edu

¹ Department of Philosophy, Chapman University, 1 University Drive, Orange, CA 92866, USA

become convinced that we should ultimately reject this principle. More than anything else, these philosophers have been persuaded of this by a class of putative counterexamples to PAP developed by Frankfurt (1969). Frankfurt-style cases, as they are called, are intended to provide possible scenarios in which agents appear to be morally responsible for their behavior even though special circumstances guaranteed that they could not have avoided doing what they did. Those interested in trying to defend PAP against this challenge from Frankfurt-style cases have two main options. The first is to argue that agents in these cases are not actually morally responsible for what they did. The second is to show that, despite efforts to rule out alternative possibilities, it was in fact possible for agents in these cases to have acted differently than they did.

This second strategy for defending PAP against Frankfurt-style attack has come to be known as the *flicker-of-freedom defense* (or the *flicker strategy*).¹ The central idea here is that, even if Frankfurt-style cases manage to successfully *restrict* the agents' alternative possibilities, they do not *eliminate* them in the way that undermining PAP would require. Although these agents do not have the wide scope of freedom we normally take ourselves to possess, they continue to be free with respect to a small but crucial set of options, and so the flickers of freedom remain. The most popular, and in my view most promising, versions of the flicker strategy focus on the fact that agents in these cases performed the relevant actions on their own, which is something they did not have to do.

For the most part, Frankfurt defenders have been willing to concede that agents in any Frankfurt-style case that is not deterministic are going to continue to have some alternative possibilities, including the possibility of refraining from performing the relevant actions on their own (and needing to be forced to perform them). Rather than seeking to eliminate this alternative, efforts have focused on arguing that this alternative possibility is insufficiently *robust* (i.e., that it is not morally significant enough to ground ascriptions of moral responsibility) and so is unable to do the work flicker theorists set out for it. In a recent issue of this journal, however, Stockdale (2022) attempts to sidestep the current debate over robustness by developing a new kind of Frankfurt-style case—one that not only makes it impossible for agents to avoid performing the actions they perform but also makes it impossible for them to avoid performing these actions *on their own*. These cases present a new and interesting challenge to this version of the flicker strategy, which, if successful, would neutralize what is probably the most promising defense of PAP against Frankfurt-style attack. My goal in this paper is to show that these versions of the flicker defense are just as effective against this new kind of case as they are against previous iterations. Although I will argue that this new kind of Frankfurt case introduced by Stockdale ultimately fails to overcome flicker strategy variants that focus on the alternative possibility of omitting to do a thing on one's own, understanding why forces us to get clearer about what it means to act on one's own and why this is morally significant. So, careful consideration of these new cases is well worth the effort.

¹ The term 'flickers of freedom' was coined by Fischer (1994: 137–47) to reference the (comparatively thin) alternatives that remain open to agents in Frankfurt-style cases.

2 Self-Inflicted Frankfurt-Style Cases

The primary goal of Frankfurt-style cases is to provide possible scenarios in which agents appear to be morally responsible for doing certain things they could not have avoided doing. Here, for instance, is one such example offered by Alfred Mele and David Robb:

Our scenario features an agent, Bob, who inhabits a world at which determinism is false. . . . At t_1 , Black initiates a certain deterministic process P in Bob's brain with the intention of thereby causing Bob to decide at t_2 (an hour later, say) to steal Ann's car. The process, which is screened off from Bob's consciousness, will deterministically culminate in Bob's deciding at t_2 to steal Ann's car unless he decides on his own at t_2 to steal it or is incapable at t_2 of making a decision (because, for example, he is dead by t_2). . . . The process is in no way sensitive to any 'sign' of what Bob will decide. As it happens, at t_2 Bob decides on his own to steal the car, on the basis of his own indeterministic deliberation about whether to steal it, and his decision has no deterministic cause. But if he had not just then decided on his own to steal it, P would have deterministically issued, at t_2 , in his deciding to steal it. Rest assured that P in no way influences the indeterministic decision-making process that actually issues in Bob's decision. (1998: 101–102; notes omitted)

The targeted actions in these cases are almost always mental acts of deciding. This is because PAP is properly understood as a thesis about direct (or non-derivative) moral responsibility and agents' responsibility for performing overt bodily actions (such as pushing someone off a cliff) derives from their responsibility for certain mental actions (such as deciding to push them off the cliff). Additionally, in nearly all cases, these targeted decisions are decisions to do something bad, for which the agents are supposed to be morally blameworthy. The reason for this is that the strength of these cases depends on others sharing the intuition that these agents are morally responsible for what they did, and empirical data confirms that we tend to have more forceful judgments of blameworthiness when agents do bad than we have of praiseworthiness when agents do good (Baumeister et al. 2001; Knobe 2003; Pizarro et al. 2003; Bostyn and Roets 2016; Guglielmo and Malle 2019).

The key element in these cases is the presence of a counterfactual intervener: some kind of agent, process, or mechanism that is poised to force the agents in these cases to make the relevant decisions if the agents do not do so on their own. By standing ready to force the agents to make the relevant decisions, these enforcement mechanisms guarantee that the agents will make the decisions they make, either on their own or as a result of being forced to do so. Crucially, however, the agents in Frankfurt cases make the targeted decisions all on their own. The enforcement mechanisms do not actually intervene; they are causally inert. Although they are present, they are irrelevant to the agents' behavior in the actual sequence of events. The fact that the ensuring conditions are causally inert and play no role in bringing about the agents' actual behavior is meant to safeguard the intuition that they are morally responsible for what they did. Because the agents in these cases acted entirely on their own, just

as they would have in ordinary circumstances, as far as their moral responsibility is concerned, it is as though the ensuring conditions were not even there.

Proponents of the flicker defense argue that, even if agents in Frankfurt-style cases are morally responsible for the decisions they made and the actions they performed, this poses no threat to PAP because these agents continued to possess alternative possibilities. According to the most popular versions of this strategy, the important alternative that remained open to agents in these cases was refraining from making the decisions they made *on their own*. Given that PAP is a thesis about direct moral responsibility, Frankfurt cases will provide us with good grounds for rejecting this principle only insofar as it was not possible for these agents to have done otherwise *than that for which they are directly morally responsible*. What these agents are directly morally responsible for, these flicker theorists maintain, is not making the decisions they made but making them *the way that they made them*—namely, on their own, without needing to be forced to make them.² Some advocates of this approach maintain that making the relevant decisions (and performing the relevant actions) on their own is the *only* thing for which agents in Frankfurt cases are morally responsible and that they are not at all responsible for making these decisions (or performing these actions) *simpliciter* (Naylor 1984; O'Connor 2000; Speak 2002; Capes and Swenson 2017). Others grant that these agents can also be morally responsible for the decisions they make (and the actions they perform) *simpliciter* but only insofar as their moral responsibility for this derives from their responsibility for making these decisions (and performing these actions) *on their own*, for which they are directly morally responsible (Robinson 2012, 2019). The central points on which they all agree are that, because it is making the relevant decisions on their own for which these agents are directly morally responsible, that is what PAP implies they must have been able to avoid. Although it was not possible for these agents not to have made the decisions they made, whether they would have to be forced to make these decisions or would instead make these decisions on their own is something that remained entirely up to them. Therefore, these cases fail to undermine the idea that moral responsibility requires the ability to do otherwise.

Contemporary advocates of Frankfurt's attack on PAP are mostly willing to grant that there are always going to be some alternatives that remain open in indeterministic Frankfurt-style cases—including the possibility of the agents needing to be forced to make the decisions they make rather than making them on their own.³ The problem with these versions of the flicker strategy, they argue, is that the alternatives

² Various arguments have been offered by proponents of this version of the flicker strategy in support of the claim that agents in Frankfurt-style cases are directly morally responsible not for making the decisions they make but for making those decisions on their own. (See Naylor 1984; Robinson 2012, 2019; Capes and Swenson 2017.) I will not rehearse those arguments here, as the truth of this claim is not required for the purposes of this paper.

³ Early Frankfurt-style cases employed signals about what actions the agents would perform if left to act on their own in order to indicate whether intervention would be necessary. These so-called “prior sign” cases faced a dilemma: Either these signs (or that which they indicated) causally determined the agents' actions or they were merely reliable but imperfect indicators of this. If the former, then these cases would be dialectically ineffective with incompatibilists, who could not be expected to share the intuition that agents in these cases are morally responsible for their actions. If the latter, then the occurrence of the signals would not strictly foreclose possibility that the agents might act differently than indicated. Following

that remain open to agents in Frankfurt-style cases are unable to do the work flicker strategists want them to do because they are insufficiently *robust* (Fischer 1994; McKenna 2003; Pereboom 2012; Sartorio 2019). The leading sentiment behind the robustness worry is the idea that, in order for it to be plausible that moral responsibility requires alternative possibilities, as PAP maintains, this must be because moral responsibility requires alternatives *of a certain sort*—namely, alternatives that are morally significant in a way that could figure into an explanation of why an agent is morally responsible for what they did (i.e., robust). If Frankfurt-style cases could show that moral responsibility does not require having alternative possibilities that are robust in this way, that would provide serious reason to doubt that moral responsibility requires alternative possibilities of any sort. This is why, rather than trying to eliminate all alternative possibilities, many prominent Frankfurt defenders have focused their attention on designing cases that succeed in eliminating all robust alternatives, even if they leave open a range of non-robust options (see Pereboom 2000, 2001, 2003, 2012; McKenna 2003; Hunt 2005). For their part, proponents of the flicker defense generally agree that defending PAP against Frankfurt-style attack requires showing that the agents in these cases continue to possess alternative possibilities that are distinctly robust in this way, but they contend that refraining from making the relevant decisions on their own is a robust alternative possibility (Speak 2002; Robinson 2012, 2014, 2019; Capes and Swenson 2017).

In a recent paper in this journal, Stockdale (2022) attempts to sidestep the current controversy surrounding the robustness issue with a new kind of indeterministic Frankfurt-style case that makes it impossible for agents to avoid making the relevant decisions on their own. Stockdale's innovation is to construct the cases such that the agents are the ones responsible for putting the ensuring conditions in place (which is why I shall refer to this new species as *self-inflicted* Frankfurt-style cases). Here is Stockdale's case (call it Anniversary):

A very forgetful self-control guru, Gary, knows that it is his twenty-fifth wedding anniversary today. Gary lives in an indeterministic world, though this does not preclude there being instances of deterministic causation in Gary's world. Knowing that he is always forgetting even the most important things as a result of getting lost in his work, Gary wants now, at a time t (before he begins work for the day), to ensure that he will decide at a later time, t_2 (after he gets off work that day), to take his wife out for dinner. Since Gary is such an exceptionally talented self-control guru, he has discovered a way to initiate a deterministic process in his brain that will ensure that a certain decision is made at a particular time of his choosing. When Gary thinks certain thoughts in a specific sequence, he can initiate a deterministic process in his own brain that will cause him to make a decision at a particular time unless his indeterministic deliberation issues in the same decision at that same time.

He then uses this technique to initiate a deterministic process (D) that will cause him to decide at time t_2 to take his wife to dinner unless his indeterministic

the publication of this problem (Kane, 1985, 1996; Widerker, 1995; Ginet 1996), cases in the literature have tended to be explicitly indeterministic.

deliberation issues in a decision at t_2 to take his wife out to dinner. Gary is so forgetful that, as soon as he starts working, he forgets about his intention to decide at t_2 to take his wife to dinner. He also forgets that he initiated and possesses D . D is screened off from the rest of Gary's consciousness and, as such, plays no role in Gary's deliberation. Now, as it just so happens, prior to t_2 Gary remembers that it is his anniversary and decides on his own at t_2 to take his wife out for dinner as a result of his indeterministic deliberation. D plays no role in his decision, as it was screened off, and Gary completely forgot about his earlier decision to take his wife to dinner and ensure the decision was made at t_2 by D . However, if Gary had not decided on his own at t_2 to take his wife out to dinner, D would have resulted in the decision at t_2 to take his wife out. (2022: 32–33)

Like virtually all other Frankfurt-type scenarios, the agent in Stockdale's case (Gary) makes a certain decision on his own, just as he would have in ordinary circumstances, and there was a special kind of enforcement mechanism that stood ready to compel the agent to make that decision unless the agent did so on his own. Unlike other cases of this kind, however, the ensuring conditions in this case were put in place by the agent himself. As a result, in the alternative sequence where the enforcement mechanism came into play and caused Gary to decide to take his wife to dinner, this would also count as a decision that Gary made on his own, according to Stockdale, because this would be a decision that Gary forced on himself, not one that was forced on him by anyone else. No matter what Gary did at the time of action, then—whether he remembered his anniversary and made the decision without any assistance or he did so as a result of the deterministic fail-safe process that he himself put in motion—he could not have avoided deciding on his own to take his wife to dinner. Insofar as this is correct, flicker strategy variants that focus on the possibility of agents omitting to make the relevant decisions on their own will be no help in defending PAP against cases of this sort. Even assuming that alternative is robust, it is not one that remains open to agents in self-inflicted Frankfurt-style cases—or so, at least, Stockdale argues. In what follows, I identify several challenges for Stockdale's case and set out to show that, though some of these problems can be avoided by making certain modifications to his example, even the strongest self-inflicted Frankfurt-style cases are not going to be able to overcome the kind of flicker defense we have been considering.

3 Designing a Better Self-Inflicted Frankfurt-Style Case

Those familiar with Frankfurt-style examples know that they tend to be rather unusual scenarios. Given that they must include some way of infallibly causing agents to make specific decisions if and only if the agents fail to make those decisions on their own, without the agent being aware of this and without having any causal impact on the agent's actual behavior, a certain amount of oddity is both inevitable and tolerable. Even by Frankfurt standards, though, Stockdale's case is pretty peculiar—peculiar in ways that threaten to hamper its ability to accomplish the work set out for it. One of the main problems with Anniversary has to do with the ensuring conditions it employs. Most Frankfurt-style cases include the presence of special devices

implanted in the agents' brains that, if necessary, will cause the agents to make the desired decisions by direct neural stimulation of the relevant parts of the brain. Of course, in typical Frankfurt scenarios, these devices are implanted by someone *other* than the agent, whereas it is crucial for the kind of case Stockdale wants to construct that this enforcement mechanism has been put in place by the agent himself, not anyone else. Rather than devising a way (and a reason) for the agent to implant such a device in himself, though, we are asked to imagine a scenario that sounds much more like sorcery than science-fiction.

In Anniversary, we are told that Gary is such a talented self-control guru that not only can he initiate a deterministic process in his brain that will yield a specific decision at an exact moment of his choosing, and not only can he create this process in a way that it will be causally preempted by (and only by) an exactly similar decision's being produced at that same exact time by the ordinary indeterministic practical deliberative processes in his brain, and not only can he ensure that this deterministic process has no causal interaction with his ordinary deliberative processes, but he can do all this *with his mind* by simply thinking certain thoughts in a particular order. We may as well have been asked to imagine a very talented magician named Harry who is capable of producing the same kind of process by incanting certain words in a specific order to cast a magical spell on himself. It is difficult to see how this would be significantly different or any more fantastical than the example Stockdale advances. In neither case would we have any explanation of the mechanism in operation—only the nebulous claim that the agent has somehow discovered a special “technique” (or spell) to bring about this effect.

Why is this a problem? After all, all that is needed to falsify PAP is a logically possible case in which an agent is directly morally responsible for doing something they could not have avoided doing; and overcoming the particular flicker strategy at issue requires only a logically possible case in which an agent is directly morally responsible for doing something that they could not have avoided doing on their own. These cases do not have to be practical or realistic to satisfy these criteria. In order to be effective, however, two things need to be intuitive to readers: that there really is something for which the agents in these cases are directly morally responsible and that this is also something these agents could not have avoided doing. In Anniversary, Stockdale needs readers to have strong intuitions both that Gary is directly morally responsible for the decision he makes on his own (at *t*₂) to take his wife to dinner and that he could not have avoided making this decision on his own.⁴ This is where the mystical nature of the ensuring condition in this case starts to work against it. The closer the operation of this backup enforcement mechanism gets to sounding like magic, the looser my grip becomes on *how* it is supposed to ensure that Gary makes that decision and the less confident I am that it is going to be able to do this without determining or otherwise causally interfering with Gary's actual behavior. Also, the more Gary starts to sound like a wizard with superhuman powers, the less sure I am

⁴ Indeed, Stockdale needs his audience to find these judgments not merely intuitive but *more intuitive* than PAP. Even if one regarded Stockdale's claims about Gary as plausible, that would not suffice to show that one should reject PAP if one considered PAP to be even more plausible.

that our everyday judgements about control and moral responsibility are going to apply to him in the same ways they apply to ordinary human beings.

The other main worry involving the particular details of Stockdale's case has to do with the agent's extraordinary forgetfulness and the roles it is meant to play in this example. One thing that Gary's forgetfulness is doing is providing his motivation for putting in place some ensuring conditions to begin with. Whereas other Frankfurt-style cases include some kind of story about the reasons some *other* agent (usually a nefarious neurosurgeon) had to implant a special device in the agent's head that would cause them to make the desired decision, Anniversary needs to include an explanation of why Gary would want to do something like this to himself. This is where Gary's fear of forgetting his anniversary and his plans to take his wife out to dinner comes into play. That is not all it is doing, however.

It is also important for Stockdale that Gary makes the decision to take his wife out to dinner at t_2 , the same time that he would make the decision if he were caused to do so by the auxiliary process he initiated at t .⁵ What makes this tricky is that Gary has clearly already decided—and so already has the intention—to take his wife out to dinner much earlier in the day, back when he enacted special measures to ensure that he followed through on his plan. Decisions are ways that we resolve practical uncertainty by actively forming intentions to act. We make decisions only when we have not yet settled what we will do (Mele 2000). If I intend to attend a friend's dinner party this weekend, I do not now need (and perhaps am even unable) to make a (further) decision to go. The same is true for Gary. As long as he is settled on taking his wife out to dinner, he need not decide that that is what he will do. For it to make sense that Gary would decide at t_2 to take his wife out to dinner that night, sometime between t and t_2 Gary needs to lose his intention to take his wife out to dinner. This is the second thing Gary's forgetfulness is supposed to accomplish in Anniversary.

Finally, it is a crucial feature of all Frankfurt-style cases that the agents are unaware of the counterfactual interveners (or counterfactually intervening mechanisms) because the guiding idea behind the design of these cases is that the special elements that guarantee the agents could not have done otherwise play no causal role in, and are irrelevant to an explanation of, the agents' actual behavior. It is this aspect of the cases that is meant to bolster the claim that the agents are morally responsible for their behavior. This is especially challenging in the kind of case Stockdale wants to construct because the central idea is that the agent himself is the one who sets up the ensuring conditions, at which point he will clearly be aware of them. So, Gary's forgetfulness is supposed to handle this as well.

We can see, then, that Gary's extraordinary forgetfulness is doing a lot of work in Anniversary. There are at least two problems with relying on this characteristic to secure these vital features of the case. First, it is not clear why Gary's forgetting (at some point between t and t_2) about his intention to take his wife out to dinner for their anniversary would result in his making a decision to do that later on at t_2 , after he remembered. As noted above, decisions are a way we can resolve practical uncer-

⁵ Presumably, this is meant to forestall versions of the flicker defense that focus on the timing of the agents' decisions in the actual and alternative sequences of these scenarios (Ginet 1996, 2002; Franklin 2011; Palmer 2011).

tainty about what we are going to do by actively forming intentions. This is not the only way we can come to have intentions, though. In cases where there is no uncertainty about what to do, we will often non-actionally acquire intentions to act without actively forming them through mental acts of deciding (Mele 2000). Normally, in a situation where I have forgotten some intention I have, unless something significant has changed that would cause me to reconsider my plans, as soon as I remember what I intended to do, I will automatically reacquire that intention. For example, imagine that I have forgotten that I am supposed to pick up my daughter from school on my way home from work today because my wife, who ordinarily picks her up, has a dental appointment. Once I remember this, I will automatically reacquire the intention to pick her up on my way home. (“Oh, right,” I would think to myself, “I’m the one who is supposed to pick up our daughter today! Whew! I’m so glad I remembered!”) I do not need to *decide* to do this, as there is no uncertainty about what to do here. It is obvious in this case that I am going to pick her up on my way home.

Something similar would seem to be true in Anniversary. During the period of time that Gary is so caught up in his work that he has forgotten about his plan to take his wife out for their anniversary, it is plausible to suppose that he no longer intends to do this and that he is not settled on doing anything in particular that night. As soon as Gary remembers this, however, it is puzzling why he would not automatically reacquire the intention to take his wife out to dinner given that there have been no relevant changes to his situation that would call for him to rethink his earlier plan. Moreover, this would still be true even if Gary had no memory of the unusual steps he took to make sure he took his wife out to dinner that night—provided he at least recalled his earlier plan to take her out. It is perhaps with this very concern in mind that we are told in Anniversary that, although Gary remembers it is his anniversary, he has no recollection *either* of his earlier intention to take his wife out to dinner *or* of the special steps he took to ensure that happened. Thus, it seems we are meant to suppose that Gary’s thought process just prior to *t*₂ went something like this: “Wait, is today the 10th? Oh, no! I completely forgot that today is our wedding anniversary! I should really do something special to show my wife that I appreciate her—and that I remembered our anniversary. What shall I do? Hmm...Should I buy her a piece of jewelry? No, I don’t have enough time to go shopping for that. I know! I’ll take her out to dinner at our favorite restaurant.” It is here that the other main worry about Gary’s forgetfulness comes to the fore.

The most serious problem with employing forgetfulness the way Stockdale does is that, in order for it to do the work it is meant to do in this case, it must be so extreme that it undermines confidence in Gary’s moral responsibility for his behavior. Consider that, even though it is of the utmost importance to him that he take his wife out for their wedding anniversary—so much so that he does everything in his considerable power to make sure he follows through on his plan—Gary forgets all about this *immediately* after he starts working. Furthermore, when he eventually does remember that today is his anniversary, he has absolutely no recollection of his intention to take his wife to dinner that night *nor* of the amped-up Jedi mind trick he used on himself *mere hours earlier* to guarantee that this happens. This is no ordinary absentmindedness. Generally, the only place one is likely to encounter this degree of memory impairment is in an elder care facility among dementia patients. As with

those suffering from dementia, this level of cognitive deficiency, though it does not necessarily preclude it, at least casts doubt on the extent to which Gary is morally responsible for his behavior. This is a critical snag for this case since its success principally depends upon readers having the strong intuition that Gary is morally responsible for his actions.

Fortunately for Stockdale, these problems are largely avoidable since there are other ways of doing the work that forgetfulness is meant to accomplish without undercutting claims about the agent's moral responsibility. To illustrate, consider another self-inflicted Frankfurt-style case, Espionage:

Larry, a scientist working at a Medtech company, has been offered a large sum of money by agents from a competitor firm to provide them with select files containing trade secrets. Retrieving these files would require accessing a restricted area of his company's building after hours and avoiding security while doing so. Although this is highly risky, Larry agrees to the arrangement because he desperately needs the money in order to settle several gambling debts that he owes to some rather unsavory characters. Larry is terrified of getting caught and is worried that he will chicken out at the last minute, so he decides to take some extraordinary measures to ensure that he follows through on his plan. One of Larry's colleagues has developed a very advanced neurological device that, when implanted in a subject's brain, can be programmed to stimulate the relevant parts of the subject's brain causing them to make a specified decision at a designated time. The way the (deterministic) device operates, it will be causally preempted by the subject's ordinary (and indeterministic) deliberative processes if at any point leading up to and including the designated time they have resulted in the subject's having the desired intention at the designated time.⁶ On top of this, Larry's company has been working on a fully automated robotic surgical system that is in the final phases of development. So, a week before the day he plans to steal the files, Larry programmed the neurological device to cause him to decide to steal the files at a certain time t late Friday night unless he makes that decision on his own. Larry then used the automated robotic surgical system to implant the device in his brain. The surgery was a success except for one minor complication: a small amount of brain tissue was damaged during the surgery which caused Larry to lose all memory having to do with his implanting a neurological device in his head. Aside from this, his cognitive functions and all other memories remain as normal. On Friday night, in the moments leading up to t , Larry was sitting in his car in the company

⁶ So, if Larry's ordinary deliberative processes result in his coming to intend to steal the files from his company and this intention persists until the designated time, this intention will causally preempt the implanted device from causing Larry to form the intention to steal the files. If his ordinary deliberative processes result in his forming the relevant intention (deciding) precisely at the designated time, that will also causally preempt the implanted device right at that moment. (For more on the plausibility of this kind of *occurrent* causal preemption, as well as the design of real-world devices that could allow for this, see Mele and Robb 2003.) In the event that Larry's ordinary deliberative processes result in his coming to intend to steal the files from his company but he loses this intention prior to the designated time, the implanted device will be causally efficacious; it will not be causally preempted by the intention that failed to persist to the designated time.

building parking lot, working up the courage to go in, when a police car raced down an adjoining street with its siren blaring. The police siren so unnerved Larry that he immediately started reconsidering his plan, unsure whether he could go through with it. After several minutes of deliberation, and without any interference from or awareness of the device implanted in his brain, Larry decided precisely at t to stick with his original plan. He entered the building, stole the files, and exited without getting caught.

There are several notable differences between *Espionage* and *Anniversary*. First, like virtually all other Frankfurt-style cases, the agent in *Espionage* performs a morally bad action for which he is meant to be not merely responsible but blameworthy. This is significant because, as previously noted, empirical evidence suggests we are more inclined to judge people blameworthy for committing bad deeds than we are to judge them praiseworthy for doing good deeds, and the success of these cases hinges on readers sharing the intuition that the agents are morally responsible for their behavior.

Second, by incorporating the same kind of surgically implanted neurological device that appears in many other Frankfurt-style cases, *Espionage* is able to avoid the worries stemming from the sort of mystical, if not outright magical, powers employed in *Anniversary*. What sets *Espionage* apart from typical cases, of course, is the fact that Larry is the one who surgically implants the special neurological device into his own brain, with the assistance of another piece of advanced technology. Although this incorporates a second science-fictional device, the automated robotic surgical system is no further beyond current technology than the described neurological device. So, it is hard to see why this should make *Espionage* any less plausible than typical Frankfurt scenarios.

Now, given that Larry intentionally took steps to ensure that he would go through with stealing the files, it seems clear that he had already decided to steal them prior to his putting the ensuring conditions in place and long before the device was set to cause him to make this decision. Insofar as it is important that Larry makes a decision to steal the files at the same time that he would have made this decision had he been caused to do so by the implanted device (time t), something will need to cause Larry to cease intending to steal the files at some point prior to t . Whereas *Anniversary* relies on Gary's extreme forgetfulness to account for this, *Espionage* accomplishes this with a temporary bout of unsettledness brought on by fear. This is the third major difference between the cases. As long as Larry is reconsidering his plan, he is no longer settled on stealing the files and he no longer intends to do so. The primary advantage of having the agent lose the pertinent intention as a result of becoming afraid rather than experiencing significant memory loss—loss so significant that it includes the agent's very important plans as well as the agent's own actions from just hours before—is that the former does not raise skeptical questions about the agent's mental acuity.⁷

The final significant difference concerns the agents' ignorance of the ensuring conditions. As discussed, it is crucial that the agents in these cases are unaware of the

⁷ This could just as easily be accomplished by having the agent reconsider his plans due to experiencing temptation instead of fear.

mechanisms that are poised to force them to make the relevant decisions if they do not do so on their own. Arranging this is particularly tricky in self-inflicted Frankfurt-style cases because the agents put these ensuring conditions in place themselves. It is hard to see an easy way around incorporating some degree of memory loss in these cases.⁸ There are different ways of accomplishing this, however, and some seem preferable to others. Whereas Gary's lack of awareness in Anniversary stems from the characteristically abysmal operation of his memory, Larry's lack of awareness in Espionage is the result of very limited, exogenously produced memory loss. Because Larry's memory loss is much more isolated and the result of external interference rather than internal deficiencies, the threat to moral responsibility in Espionage seems less significant than in Anniversary. At least, that is the hope.

Perhaps some will worry that these modifications are insufficient to completely avoid the worries raised about Stockdale's example in the previous section. That is fine. My immediate goal is to show that some of the problems facing Anniversary can be avoided and to produce the strongest version of a self-inflicted Frankfurt-style case that I can. I do not aim to argue that Espionage is successful. On the contrary, in what follows I set out to demonstrate that not even the best self-inflicted Frankfurt-style cases are going to be able to overcome the on-your-own versions of the flicker defense.

4 Acting on One's Own

As with all indeterministic scenarios of this sort, agents in self-inflicted Frankfurt-style cases continue to retain a limited range of alternative possibilities. In Espionage, for instance, Larry makes the decision to steal the files without needing to be forced to do so, but it was also possible for him not to have done this (in which case he would have been forced to make that decision by the device implanted in his brain). The linchpin of Stockdale's argument is the claim that, even if Larry had been forced to make the decision to steal the files by the mechanism he himself put in place, that still would have been a decision he made *on his own*. If that claim stands, then, even if one were to grant the flicker theorist view that what agents in these cases are directly morally responsible for is making the relevant decisions on their own, that will not be adequate to defend PAP against self-inflicted Frankfurt-style cases because making those decision on their own was not something they could have avoided doing (at least, not at the crucial time leading right up to their decisions).

4.1 Stockdale's Conception

What should we make of the idea that decisions agents are forced to make by conditions they themselves put in place still qualify as decisions made on their own (in the

⁸ One possible way to obviate the need for some level of memory loss could be for the agent to somehow set up the ensuring conditions without any awareness of doing this. Whether this could be done without making the cases substantially more bizarre than they already are, and so further endangering their intuitiveness, is unclear.

sense at issue in this debate)? As Stockdale points out, there are different ways of characterizing what it means for an agent to do a thing on his own that have appeared in the Frankfurt literature. Whereas I (Robinson 2019: 217 n. 14) have used the locution to mean that an agent did a thing “without being causally forced or interfered with (in the way that he would have been in the alternative scenario in a Frankfurt-style case),” Capes (2014: 428) says it indicates that the agent’s doing it “was not a result of external coercion or force by the likes of [the counterfactual intervener’s] device,” and Mele (1996: 126) uses it to mean that “no one else” made the agent do it. Here we have three potentially different ways of thinking about what it is to do a thing on one’s own:

- O1. To do a thing on one’s own is to do it without being forced to do it.
- O2. To do a thing on one’s own is to do it without being forced to do it by anything external.
- O3. To do a thing on one’s own is to do it without being forced to do it by someone else.

If O1 is the right way to think about what it means to act on one’s own, then self-inflicted Frankfurt-style cases will not be invulnerable to the versions of the flicker defense under consideration since the decisions agents make in the alternative sequences of these cases would not count as decisions they make on their own. Indeed, it is precisely because I was operating with the notion of acting on one’s own described in O1 that I argued that the alternative possibility of refraining from making the relevant decisions on their own is ineliminable from indeterministic Frankfurt cases. As I put it elsewhere, “There is no way to guarantee that an agent will make a certain decision on his own because it is impossible (indeed, incoherent) to *force* an agent to do something *on his own* (that is, without being forced)” (Robinson 2019: 229).

Perhaps unsurprisingly, then, Stockdale contends that we should not accept the characterization of doing a thing on one’s own laid out in O1 and should instead opt for one of the other two definitions. He does not say which of the alternative conceptions he views as correct (O2 or O3), only that we should reject O1. Possibly, this is because he views O2 and O3 as being similar (Stockdale 2022: 34). This also might be owing to the fact that he does not see this as mattering for the purposes of his argument since, in his view, both accounts imply that decisions agents are forced to make by conditions they themselves put in place are decisions they made on their own. Speaking of his Anniversary case, he writes:

Gary’s deterministic process (*D*) producing a decision would not constitute a decision that anyone else made him perform, nor was it a result of external coercion or force. *D* originated from Gary himself, such that no one else was involved in its creation or at any time between its creation and *t*₂. At no time was it external to him, and neither was it the result of external coercion or force. No matter whether *D* or Gary’s indeterministic mental processes culminate in Gary’s decision to take his wife to dinner, according to the usages and defini-

tions discussed above [what I have labeled O2 and O3], the decision will be one he made on his own. (2022: 34)

4.2 Stockdale's First Argument: Usage in the Literature

Stockdale offers two reasons in support of thinking that we should reject O1 in favor of O2 and O3. First, he contends that this is the way Frankfurt used the phrase and intended it to be understood when introducing his original case. Citing Frankfurt's (1969) claim that "whether [Jones] finally acts on his own or as a result of Black's intervention, he performs the same action," Stockdale writes,

This quotation...illustrate[s] that 'on his own' was meant to pick out the sequence in which there is no manipulation or coercion by Black, and many philosophers have followed Frankfurt's lead and used the phrase to pick out an action performed in the absence of force by another agent. (2022: 34)

Stockdale references Mele as someone he views as following Frankfurt's lead in taking the claim that an agent did a thing on their own to mean that they did it without being forced to do it by anyone else (as stated in O3), and he sees Capes as thinking something similar (with O2). He also points to Naylor (1984), Fischer (1994), and McKenna (1997) as further examples of those who discuss acting on one's own in the same manner as Frankfurt (Stockdale 2022: 34).

There are a few points to make about this first line of argument. To begin with, aside from Mele, it is not at all clear that these cited authors are operating with the same kind of conception of acting on one's own that Stockdale has in mind (one that applies to cases in which an agent's action is forced by conditions they themselves have put in place).⁹ Consider Capes's (2014) claim that to say that an agent did a thing on their own indicates that it "was not a result of external coercion or force by the likes of [the counterfactual intervener's] device" (428). Stockdale appears to read Capes as endorsing something like O2. But O2 itself is ambiguous, as there are different things one could mean in saying that an action or decision is the result of external force. One thing this could mean is that it was forced by something external to one's body or, perhaps more specifically, to one's brain. Pettit (2005), on the other hand, suggests another interpretation when claiming that "there is a difference between doing something 'on one's own' and being coerced into doing something by a mechanism external to one's normal decision making process" (309). So, here we have two ways of resolving the ambiguity in O2:

O2a. To do a thing on one's own is to do it without being forced to do it by anything external to one's brain.

⁹ Even in Mele's case, he states, "The expressions 'on her own' and 'not on her own', as used here, are technical expressions; I am not attempting to capture ordinary usage" (1996: 126). So, even though he is certainly operating with the definition in O3, it is not clear what kind of support this provides for thinking this is the way Frankfurt and others in this literature have been conceiving of it since Mele is offering a stipulative definition that is not intended to line up with the way others have defined it.

O2b. To do a thing on one's own is to do it without being forced to do it by anything external to one's ordinary decision-making processes.

Even if agents in self-inflicted Frankfurt-style cases could not avoid making the targeted decisions on their own in the sense specified in O2a, they certainly could have avoided making them on their own in the sense spelled out in O2b. So, the success of these cases requires that we go along with Stockdale in rejecting not only O1 but also O2b as characterizing the relevant sense of acting on one's own. Given Stockdale's plain view that Gary's decision in the alternative sequence of Anniversary would *not* have been the result of external force or coercion, it is clear that it is O2a, rather than O2b, that captures the way Stockdale is thinking about external force. It is far less clear, however, whether this is the same conception Capes has in mind or whether he is thinking about this closer to the way Pettit (2005) describes. At best, Capes's (2014) statement is neutral with respect to O2a and O2b.

The passages Stockdale cites from the other authors are similarly ambiguous with respect to O1, O2a, O2b, and O3. Since this is one of the two main reasons Stockdale offers in support of his conception of doing a thing on one's own, which is crucial to the success of his argument, it is worth briefly demonstrating that this is the case. When discussing an example involving a "child who tells the truth when he would have been forced to do so anyway," Naylor says the child "clearly deserves moral credit for telling the truth *on his own*—i.e., without being forced to tell the truth" (1984: 251). Describing the earliest versions of this flicker approach, Fischer writes that "we mean by 'on his own' at least in part 'not as a result of some weird intervention such as that of Black'" (1994: 139). According to McKenna, "In the Frankfurt-type cases the alternatives are, either doing what one does of one's own intention, or being coerced into performing the same kind of action against one's will" (1997: 75). None of these passages strongly suggest that their authors understand doing a thing on one's own in the ways laid out in O2a and O3 as opposed to those stated in O1 and O2b. If anything, they sound closer to O1 and the way Robinson (2019) describes it.

The same is true of Frankfurt's statements in his original paper. So, pointing to the way that Frankfurt and others have used the expression "on his own" is not going to provide a strong reason for defining it in one of the ways needed for Stockdale's argument to go through. Indeed, insofar as one regards PAP as plausible and is interested in defending it against the challenge posed by (self-inflicted) Frankfurt-style cases, Stockdale's argument will have provided a considerable reason *against* defining it in this way.

4.3 Reasons to Resist Stockdale's Conception

There are also independent reasons for thinking that we should reject the characterizations of doing a thing on one's own stated in O2a and O3. To see this, it will be helpful to consider some examples to bring out the more problematic implications of these conceptions. Begin, then, by imagining an eccentric (and somewhat sadistic) billionaire who, inspired by Kavka's (1983) toxin puzzle, decides to hold a contest involving a non-lethal toxin that, if consumed, would make a person very ill for a day but have no lasting negative effects. Whereas participants in Kavka's example had

only to *intend* to drink the toxin by a certain time in order to win the million-dollar prize, participants in this version must actually drink the toxin to win. Also unlike Kavka's version, there are no prohibitions on enlisting outside assistance to help one secure the prize money. Suppose I have been selected to participate in this contest and that, although I desperately want the prize money and am willing to endure the temporary sickness to obtain it, I am worried that I might chicken out when the critical moment comes (especially if I were to observe other participants suffering the ill effects of drinking the toxin). Therefore, I have hired four quick and powerful bodybuilders (to be paid out of my winnings) to hold me down and force me to drink the toxin if I hesitate too long—say, more than a minute. Now, if I were to wait longer than that and these bodybuilders were to overpower me and forcibly pour the toxin down my throat—despite my kicking and screaming and protesting that I had changed my mind—what should we say about such a case? Would my drinking the toxin in this scenario be accurately described as something I did *on my own*? Hopefully, readers will share the intuition that there is no significant sense in which I consumed the toxin on my own under such circumstances. On the contrary, being physically overpowered and compelled to do a thing is about as close to the opposite of doing it on one's own as seems possible. It is true, of course, that the bodybuilders were only doing what I hired them to do. But, while it is plausible to suppose that my making arrangements to be forced to drink the toxin is something I did on my own, it is implausible to think that I drink the toxin on my own when I am physically forced to do so. Instead, we should say that I brought it about on my own that I was forced to drink the toxin. That is, on my own I brought it about that I drank the toxin not on my own.

Neither O2a nor O3 imply that I drink the toxin on my own when I am forcibly made to do so by the bodybuilders I engaged for that purpose since my drinking the toxin in that scenario would be something that I was both forced to do by something external to my brain and forced to do by another person. So, this case poses no problem in itself. The trouble comes when we consider a slight variant of this example. Suppose that, instead of hiring bodybuilders, I borrowed several humanoid robots from the engineering lab on campus and programmed them to grab me and force me to down the toxin if I faltered for more than a minute. Should this modification to the example have any impact on our view about whether drinking the toxin is a thing I do on my own? It is hard to see why. After all, if I am forced to do something by someone else, it clearly makes no difference which particular person forced me to do it. (No one in the Frankfurt literature has proposed that doing a thing on one's own means doing it without being forced to do it *by Black*, for instance.) But, then, if it does not matter which particular person forces me to do a thing—whether it is Black, Brown, Gray, or whomever—it is unclear why it would matter whether what forces me to do it is a person at all (or whether it was programmed or initiated by a person).¹⁰

¹⁰ Mele (2006) makes a similar point in the course of laying out his zygote argument when he suggests that replacing the manipulator in a manipulation case with blind natural forces should have no impact on our view of the manipulated agent's actions and the extent to which the manipulated agent performs them freely and is morally responsible for them (190; see also 198, n. 16).

According to O3, however, this is the very feature that distinguishes actions done on one's own from those not done on one's own. O3 implies that I drink the toxin on my own when I am forced to do so by robots I have programmed—because I am not forced to drink it *by someone else*—but I do not drink it on my own when I am forced to do so by bodybuilders I have hired (and, presumably, not when I am forced to do so by robots that have been programmed by someone else). Seemingly, though, in a situation where robots overpower me and force me to swallow the toxin, it makes no difference whether the robots were programmed to do that by another person or they randomly popped into existence and just so happened to do that. The effect on me is the same. Either way, I do not drink the toxin on my own. Without a compelling reason for thinking that it matters for my doing a thing on my own whether I am forced to do it by something that is a(nother) person, it seems that, whatever we say about my drinking the toxin when forced to do so by bodybuilders I hired, we should say the same thing about my drinking the toxin as a result of being forced to do so by robots I programmed (as well as my drinking the toxin when made to do so by robots functioning at random that were not programmed by anyone). Insofar as that is correct, we have good grounds for rejecting the definition of acting on one's own laid out in O3.

This is not the only strange implication of O3. Suppose that you had been observing the contest and had watched as several robots chased me down, restrained me, pried open my mouth, poured the toxin down my throat, and held my mouth shut until I swallowed it. And now imagine that someone who was not present at the contest later asked you whether I had drunk the toxin on my own. What would you say? Would you say that I had not drunk the toxin on my own, that I had been forced to drink it by robots that overpowered me? Although this seems like a perfectly reasonable (and perhaps the most natural) thing to say in the circumstances, O3 implies that you are in no position to answer the question because you do not know who, if anyone, programmed the robots to do that. In order to know whether I drank the toxin on my own, according to O3, you would need to know whether someone else programmed the robots. If so, then I did not drink the toxin on my own. But if *I* programmed the robots to do that, or if *nobody* programmed the robots to do that and their behavior was simply the result of random short-circuiting, O3 would count my drinking the toxin as something I did on my own.

The examples just considered all involve my being forced to do a thing by objects external to my brain. So, they are not going to pose a problem for O2a, which would not count any of those as things I did on my own. But O2a is problematic too. For, just as it is difficult to see why it should matter whether an external force that compels an agent to do a thing is a person or not, it is also unclear why it matters whether the source of the compulsion is internal or external to one's brain. In a case where a scientist who wanted to watch me suffer the ill effects of the toxin programmed robots to overpower me and force me to drink it, which they succeed in doing, it is uncontroversial that my drinking the toxin would not be something I did on my own. What sense does it make to maintain that, if a scientist were to program robots to force me to drink the toxin, that would not be something I did on my own but that, if a scientist were to implant a device in my brain that made me decide to drink the toxin, *then* my drinking the toxin would be something I did on my own? If a scientist sets out to force

me to drink a toxin, why does it matter whether he does this by programming robots to overpower me or by implanting a device in my brain? The physical location of the compelling force seems irrelevant. All that matters is that it is forcing, compelling, coercing my behavior.

Now, none of these cases are problematic for the conceptions of doing a thing on one's own found in O1 and O2b. Because they all involve my being forced to drink the toxin—and my being forced to drink it by things external to my ordinary decision-making processes—both O1 and O2b correctly imply that I do not drink the toxin on my own in any of these scenarios. The trouble for Stockdale's argument is that these conceptions do not support—indeed, they undermine—the crucial claim that agents in self-inflicted Frankfurt-style cases could not have avoided making the relevant decisions on their own. It is trivially true that when these agents are forced to make the targeted decisions by the devices they have put in place (or processes they have initiated)—as they are in the alternative sequences of self-inflicted Frankfurt-style cases—making those decisions is something they are *forced* to do. So, making those decisions will not count as things they do on their own according to O1. It is also clear that that the special, counterfactually intervening enforcement mechanisms agents employ in these cases are not part of their ordinary decision-making processes. Thus, the decisions they are forced to make by these mechanisms will not qualify as ones made on their own by O2b either.

4.4 Stockdale's Second Argument: Diachronic Self-Control

There is one other line of argument that Stockdale offers in support of the claim that the decisions agents make in the alternative sequences of self-inflicted Frankfurt-style cases—wherein they are forced to make those decisions by the ensuring conditions they themselves put in place—are decisions they make on their own. This argument appeals to Kennett and Smith's (1996) understanding of diachronic self-control and the idea that we “can exercise control diachronically, at an earlier time, by so arranging the circumstances of action at the later time as to remove the possibility of our then losing control” (68). Stockdale argues that the alternative sequences of self-inflicted Frankfurt-style cases are instances of diachronic self-control and that agents can act on their own even when their actions are the result of prior exercises of diachronic self-control. Of course, even if we were to grant both that the alternative sequences of self-inflicted Frankfurt-style cases are instances of diachronic self-control and that *some* instances of diachronic self-control result in actions that still count as things the agents do on their own, that would not be sufficient to demonstrate that (any of) the agents in the alternative sequences of self-inflicted Frankfurt-style cases make the relevant decisions on their own. That would follow only if we were to grant that *all* behaviors that result from exercises of diachronic self-control count as things the agents do on their own. Even Stockdale refrains from suggesting that this is true (and rightly so).

Instead, Stockdale proceeds to offer an argument from analogy based on the following case:

Betsy wants to work on her dissertation from 9 a.m. – 5 p.m. every day. Unfortunately, she gets distracted easily and will spend large portions of the day watching TV instead of getting work done. She knows herself well enough to know that if there is a functioning TV in her apartment, then she will be unable to resist her urge to procrastinate by watching it. One day she finally has enough, so Betsy wants now to make sure that she will decide at 9 a.m. tomorrow to work on her dissertation. To guarantee that she will decide tomorrow to work on her dissertation and no longer get distracted by the TV, she breaks her TV. The next day at 9 a.m. Betsy feels a strong desire to watch TV instead of work, but the TV does not work. If the TV did work, she would cave to her desire to watch it, but since it does not, she decides to work on her dissertation. (2022: 38)

Stockdale contends that “there are no relevant differences between normal diachronic self-control cases and the outcome in my [Frankfurt-style case] in which the decision is made as a result of the deterministic process” (39). In his view, both cases involve agents taking steps to ensure that they behave a certain way in the future and their making decisions that cannot be attributed to anyone else. As a result, he concludes, “If it is right to think that Betsy decided on her own to work on her dissertation, it is similarly true that Gary decided on his own at t_2 to take his wife out to dinner in the scenario in which his decision was produced by D ” (38).

The problem with this argument is that there is a key difference between these cases which undermines the analogy on which the argument rests. Betsy is not *forced* to decide to work on her dissertation. No doubt, by removing a potential source of temptation, Betsy’s earlier action of breaking the television made it easier for her to make the decision to work on her dissertation (and so increased the likelihood that she would do so). But the absence of a functioning television does not actually force or compel her to decide to work on her dissertation, nor does it preclude her from deciding to do something else. She could have decided to read a book, take a walk, go back to bed, drive to the store to buy a new television, or any number of activities besides deciding to work on her dissertation. In the alternative sequences of self-inflicted Frankfurt-style cases, however, the mechanisms the agents put in place physically *compel* them to make the relevant decisions and preclude them from doing anything else. Indeed, that is their purpose.

Contrary to Stockdale’s contention, then, appealing to diachronic self-control does not seem to provide support for the idea that agents in the alternative sequences of self-inflicted Frankfurt-style cases make the targeted decisions on their own. There may well be some scenarios, such as Stockdale’s example involving Betsy, in which an agent’s decisions are partly the result of previous exercises of diachronic self-control and, intuitively, are also decisions they make on their own. But these cases are importantly dissimilar to the decisions agents are forced to make in the alternative sequences of cases like Anniversary and Espionage. There is a significant difference between eliminating potential sources of temptation and arranging to be physically compelled to act in a certain way (whether by people, robots, brain implants, or otherwise). When Odysseus orders his men to lash him to the mast of the ship and not to unbind him (but, rather, to bind him tighter) should he demand to be released before

they have passed out of range of the sirens singing their song, it may well be that he is exercising a form of diachronic self-control. Even so, it is implausible to suggest that Odysseus remains on the ship *on his own* when, despite his best efforts to the contrary, he is kept on the ship only by his crew and the ropes that they use to forcibly restrain him. The same goes for the examples discussed earlier in which I hire bodybuilders and program robots to run me down and force me to drink the toxin in the event that I waver due to fear of becoming very ill. Insofar as they involve agents making arrangements to be forced to make the desired decisions, rather than simply removing sources of temptation to act contrarily, the alternative sequences of self-inflicted Frankfurt-style cases are much more analogous to these latter scenarios than they are to the example involving Betsy breaking her television.

5 Conclusion

Self-inflicted Frankfurt-style cases offer a novel and clever challenge to the most promising versions of the flicker defense of PAP. Ultimately, however, they are unable to eliminate the alternative possibility they are meant to rule out, and so they remain just as vulnerable to this defense as are typical Frankfurt-style cases. As I have shown in this paper, the fundamental flaw in Stockdale's argument involves the conception of acting on one's own that lies at its core. Neither of the two main reasons he offers—not the way this phrase has been used and characterized by prominent participants in the debate nor consideration of cases involving diachronic self-control—provide strong support for thinking that the decisions agents are forced to make in the alternative sequences of self-inflicted Frankfurt-style cases are decisions they make on their own. Moreover, the conceptions of acting on one's own that Stockdale's argument requires in order to succeed have a number of problematic implications. They imply that I perform an action on my own when I program robots to overpower me and make me do it but not when I hire bodybuilders to force me to do the same. They imply that I perform an action on my own when I am forced to perform it by powerful robots that are randomly malfunctioning but not when forced to do so by robots that have been programmed by someone else. They imply that I do not make a decision on my own when caused to do so by a device that someone else has implanted in my brain but that I do make that decision on my own when forced to do so by a neural device I have implanted myself (or by a small electromagnetic field, located entirely within the boundary of my brain, that appears without being caused to by any agent but that effectively functions just like the brain implant, compelling me to make that decision by directly stimulating the relevant portions of my brain). Thus, while Stockdale's paper does a service to this corner of the philosophical debate concerning free will and moral responsibility by drawing attention to the need for greater clarity around what it means in this context to do a thing on one's own, the particular characterizations that he endorses are untenable.

There is one final point worth emphasizing here. Whereas Stockdale's argument succeeds only insofar as others are willing to accept his favored characterization(s) of what it is to act on one's own, the versions of the flicker strategy that these cases are meant to overcome are not similarly dependent on how one defines any particu-

lar term. This is because it is not fundamentally a semantic point for which they are arguing. Their central argument can be made just as easily without the use of this specific phrase. Proponents of these versions of the flicker defense argue that agents in Frankfurt-style cases are (directly) morally responsible not merely for making the decisions they make but for making them *the way that they make them*. Although there were mechanisms in place ready to compel the agents to make the pertinent decisions, these ensuring conditions were unnecessary because the agents acted in the desired ways without any awareness of or causal interference from the unusual contrivances at hand. They made their decisions without needing to be forced to make them. *That* is what the agents are (directly) morally responsible for, according to these flicker theorists. And there is a real, substantive difference between doing a thing only because you were forced to do it and doing it without needing to be forced. Indeed, it is an essential feature of Frankfurt-style cases that the agents do what they do without needing to be forced to do so, and it is crucial to securing the intuition that the agents are (directly) morally responsible for their behavior.

One perfectly natural way to express this substantive claim—that it is making the relevant decisions *without needing to be forced to do so* for which the agents in these cases are (directly) morally responsible—is to say that the agents are (directly) morally responsible for making these decisions *on their own*.¹¹ Again, however, the very same point can be made without employing this particular locution. Suppose that these flicker theorists were to set aside the phrase ‘on their own’ and lay out their arguments in terms of agents being (directly) morally responsible for making the decisions they make without needing to be forced to do so. Where would that leave Stockdale’s argument? Even in self-inflicted Frankfurt-style cases, it is clear that the agents could have refrained from making the targeted decisions without being forced to make them (by the mechanisms they themselves had put in place). That the potential threat from self-inflicted Frankfurt-style cases is so yoked to the precise wording used to describe the relevant alternatives and fades so quickly once different language is used to pick them out seems to indicate that Stockdale’s challenge consists primarily in exploiting a potential semantic ambiguity rather than in identifying a deeper weakness in these versions of the flicker defense.

Funding Open access funding provided by SCEL, Statewide California Electronic Library Consortium. The author has no relevant financial or non-financial interests to declare.

Open access funding provided by SCEL, Statewide California Electronic Library Consortium

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use

¹¹ Indeed, to underscore just how natural this is, it is worth observing that this is precisely how Stockdale himself uses the phrase when offering his own self-inflicted Frankfurt-style case, Anniversary. When Stockdale wants to communicate to readers that Gary makes the desired decision without being forced to do so by the ensuring conditions that Gary himself put in place, he writes: “Now, as it just so happens, prior to t_2 Gary remembers that it is his anniversary and decides *on his own* at t_2 to take his wife out for dinner as a result of his indeterministic deliberation” (Stockdale 2022: 33; emphasis mine).

is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Baumeister, R.F., E. Bratslavsky, C. Finkenauer, and K.D. Vohs. 2001. Bad is stronger than good. *Review of General Psychology* 5(4): 323–370. <https://doi.org/10.1037/1089-2680.5.4.323>.
- Bostyn, D. H., and A. Roets. 2016. The morality of action: the asymmetry between judgments of praise and blame in the action-omission effect. *Journal of Experimental Social Psychology* 63: 19–25. <https://doi.org/10.1016/j.jesp.2015.11.005>.
- Capes, J. A. 2014. The flicker of freedom: a reply to stump. *The Journal of Ethics* 18: 427–435. <https://doi.org/10.1007/s10892-014-9181-2>.
- Capes, J. A., and P. Swenson. 2017. Frankfurt cases: the fine-grained response revisited. *Philosophical Studies* 174: 967–981. <https://doi.org/10.1007/s11098-016-0726-z>.
- Fischer, J. M. 1994. *The Metaphysics of Free Will*. Blackwell Publishers Inc.
- Frankfurt, H. 1969. Alternate possibilities and moral responsibility. *Journal of Philosophy* 66(23): 829–839. <https://doi.org/10.2307/2023833>.
- Franklin, C. 2011. Neo-frankfurtians and buffer cases: the new challenge to the Principle of Alternative possibilities. *Philosophical Studies* 152: 189–207. <https://doi.org/10.1007/s11098-009-9472-9>.
- Ginet, C. 1996. In defense of the Principle of Alternative possibilities: why I don't find Frankfurt's arguments convincing. *Philosophical Perspectives* 10: 403–417. <https://doi.org/10.2307/2216254>.
- Ginet, C. 2002. Review of living without Free Will. *Journal of Ethics* 6: 305–309. <https://doi.org/10.1023/A:1019569214450>.
- Guglielmo, S., and B. F. Malle. 2019. Asymmetric morality: blame is more differentiated and more extreme than praise. *PLOS ONE* 14(3)Article e0213544. <https://doi.org/10.1371/journal.pone.0213544>.
- Hunt, D. 2005. Moral responsibility and buffered alternatives. *Midwest Studies in Philosophy* 29(1): 126–145. <https://doi.org/10.1111/j.1475-4975.2005.00109.x>.
- Kane, R. 1985. *Free will and values*. State University of New York Press.
- Kane, R. 1996. *The significance of free will*. New York: Oxford University Press.
- Kavka, G. 1983. The toxin puzzle. *Analysis* 43(1): 33–36. <https://doi.org/10.2307/3327802>.
- Kennett, J., and M. Smith. 1996. Frog and toad lose control. *Analysis* 56: 63–73. <https://doi.org/10.1093/analysis/56.2.63>.
- Knobe, J. 2003. Intentional action and side effects in ordinary language. *Analysis* 63(279): 190–194. <https://doi.org/10.1111/1467-8284.00419>.
- McKenna, M. 1997. Alternative possibilities and the failure of the counterexample strategy. *Journal of Social Philosophy* 28(3): 71–85. <https://doi.org/10.1111/j.1467-9833.1997.tb00388.x>.
- McKenna, M. 2003. Robustness, control, and the demand for morally significant alternatives: Frankfurt examples with oodles and oodles of alternatives. In D. Widerker & M. McKenna (Eds.), *Moral responsibility and alternative possibilities: Essays on the importance of alternative possibilities* (pp. 201–218). Ashgate Publishing Company. <https://doi.org/10.4324/9781315199924>.
- Mele, A. 2006. *Free Will and Luck*. Oxford University Press.
- Mele, A. R. 1996. Soft Libertarianism and Frankfurt-style scenarios. *Philosophical Topics* 24(2): 123–141. <https://doi.org/10.5840/philtopics199624220>.
- Mele, A. R. 2000. Deciding to act. *Philosophical Studies* 100: 81–108. <https://doi.org/10.1023/A:1018639622382>.
- Mele, A. R., and D. Robb. 1998. Rescuing Frankfurt-style cases. *The Philosophical Review* 107: 97–112. <https://doi.org/10.2307/2998316>.
- Mele, A.R., and D. Robb. 2003. Bbs, magnets, and seesaws: The metaphysics of Frankfurt-style cases. In *Moral responsibility and alternative possibilities: Essays on the importance of alternative possibilities*, eds. D. Widerker and M. McKenna, 127–138. Ashgate Publishing Company. <https://doi.org/10.4324/9781315199924>.
- Naylor, M. 1984. Frankfurt on the principle of alternate possibilities. *Philosophical Studies* 46: 249–258. <https://doi.org/10.1007/BF00373108>.

- O'Connor, T. 2000. *Persons and causes: the metaphysics of free will*. Oxford University Press. <https://doi.org/10.1093/019515374X.001.0001>.
- Palmer, D. 2011. Pereboom on the Frankfurt cases. *Philosophical Studies* 153: 261–272. <https://doi.org/10.1007/s11098-009-9489-0>.
- Pereboom, D. 2000. Alternative possibilities and causal histories. *Philosophical Perspectives* 14: 119–137. <https://doi.org/10.1111/0029-4624.34.s14.7>.
- Pereboom, D. 2001. *Living without free will*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511498824>.
- Pereboom, D. 2003. Source incompatibilism and alternative possibilities. In *Moral responsibility and alternative possibilities: Essays on the importance of alternative possibilities*, eds. D. Widerker and M. McKenna, 127–138. Ashgate Publishing Company. <https://doi.org/10.4324/9781315199924>.
- Pereboom, D. 2012. Frankfurt examples, derivative responsibility, and the timing objection. *Philosophical Issues* 22: 298–315. <https://doi.org/10.1111/j.1533-6077.2012.00231.x>.
- Pettit, G. 2005. Moral responsibility and the ability to do otherwise. *Journal of Philosophical Research* 30: 303–319. <https://doi.org/10.5840/jpr20053040>.
- Pizarro, D., E. Uhlmann, and P. Salovey. 2003. Asymmetry in judgments of moral blame and praise: the role of perceived metadesires. *Psychological Science* 14(3): 267–272. <https://doi.org/10.1111/1467-9280.03433>.
- Robinson, M. 2012. Modified Frankfurt-type counterexamples and flickers of freedom. *Philosophical Studies* 157: 177–194. <https://doi.org/10.1007/s11098-010-9631-z>.
- Robinson, M. 2014. The Limits of limited-blockage Frankfurt-style cases. *Philosophical Studies* 160: 429–446. <https://doi.org/10.1007/s11098-013-0190-y>.
- Robinson, M. 2019. Robust flickers of freedom. *Social Philosophy and Policy* 36(1): 211–233. <https://doi.org/10.1017/S0265052519000244>.
- Sartorio, C. 2019. Flickers of freedom and moral luck. *Midwest Studies in Philosophy* 43: 93–105. <https://doi.org/10.1111/misp.12106>.
- Speak, D. 2002. Fanning the flickers of freedom. *American Philosophical Quarterly* 39(1): 91–105.
- Stockdale, B. 2022. Moral responsibility, alternative possibilities, and acting on one's own. *The Journal of Ethics* 26(1): 27–40. <https://doi.org/10.1007/s10892-020-09353-z>.
- Widerker, D. 1995. Libertarianism and Frankfurt's attack on the principle of alternative possibilities. *Philosophical Review* 104: 247–261. <https://doi.org/10.2307/2185979>.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.