

ELEONORE STUMP

ALTERNATIVE POSSIBILITIES AND MORAL RESPONSIBILITY:  
THE FLICKER OF FREEDOM

(Received 5 February 1999; accepted in revised form 28 April 1999)

**ABSTRACT.** Some defenders of the principle of alternative possibilities (PAP) have responded to the challenge of Frankfurt-style counterexamples (FSCs) to PAP by arguing that there remains a “flicker of freedom” – that is, an alternative possibility for action – left to the agent in FSCs. I argue that the flicker of freedom strategy is unsuccessful. The strategy requires the supposition that doing an act-on-one’s-own is itself an action of sorts. I argue that either this supposition is confused and leads to counter-intuitive results; or, if the supposition is acceptable, then it is possible to use it to construct a FSC in which there is no flicker of freedom at all. Either way, the flicker of freedom strategy is ineffective against FSCs. Since the flicker of freedom strategy is arguably the best defense of PAP, I conclude that FSCs are successful in showing that PAP is false. An agent can act with moral responsibility without having alternative possibilities available to her.

**KEY WORDS:** alternative possibilities, causal determinism, compatibilism, flicker of freedom, Frankfurt-style counterexamples, free will, incompatibilism, indeterminism, libertarianism, moral responsibility

INTRODUCTION

That moral responsibility entails indeterminism<sup>1</sup> is a thesis which is not attractive to many philosophers, but it is a consequence of two very plausible intuitions, Peter van Inwagen says.<sup>2</sup> One intuition is that moral responsibility requires the ability to act otherwise than one in fact does;<sup>3</sup>

---

<sup>1</sup> For the purposes of this paper, I will understand an indeterministic act *A* on the part of some agent *S* as an act that *S* does without *S*’s having been causally determined to do *A* by a cause which is part of an unbroken causal chain that originates in something other than *S*’s own intellect and will. I don’t know what exactly van Inwagen understands an indeterministic act to be, but since he plainly wants it to be an act of the sort that can’t occur in a fully deterministic world, perhaps my understanding and his at least overlap on the central point.

<sup>2</sup> Peter van Inwagen, “Fischer on Moral Responsibility,” *The Philosophical Quarterly* 47 (1997), p. 374.

<sup>3</sup> Van Inwagen uses “free will” as synonymous with “the ability to act otherwise than one in fact does”; but this usage seems to me confusing in the current discussion. For the sake of clarity, I will use “the ability to act otherwise” where van Inwagen has “free will.”



the second is that agents cannot have the ability to act otherwise in a fully deterministic world.<sup>4</sup> Contrary to what he expects of those who disagree with him, I share van Inwagen's view that moral responsibility entails indeterminism, but not both the intuitions on which he bases that view.

The view that moral responsibility entails indeterminism is, as van Inwagen says, controversial; and it is not part of my purpose in this paper to try to defend it.<sup>5</sup> Instead, I want to argue against the first of van Inwagen's intuitions, namely, that moral responsibility requires the ability to do otherwise. Since, however, I agree with van Inwagen that moral responsibility does entail indeterminism, I will also argue that an agent who acts indeterministically need not have alternative possibilities for acting open to her. I share the second of van Inwagen's intuitions, that an agent's being causally determined is incompatible with her having the ability to act otherwise; but I will argue that an agent's not being causally determined, her acting indeterministically, isn't sufficient for her to have that ability, and so it isn't necessary that an agent who acts indeterministically have alternative possibilities for her action.

People who share van Inwagen's view suppose that the principle of alternative possibilities (PAP) is true. PAP has many different formulations, but they all try to capture the first of van Inwagen's two intuitions, namely, that

---

<sup>4</sup> For present purposes, I take a fully deterministic world as one in which everything that happens is fully necessitated by antecedent causes; most contemporary versions of causal determinism also include the presupposition that the necessitating antecedent causes are at the microphysical level. As I explain indeterministic acts in footnote 1, there can be no indeterministic acts in a fully deterministic world. Strictly speaking, however, the world could fail to be fully deterministic without its containing indeterministic human acts of the sort described in footnote 1, since the world could fail to be fully deterministic in virtue of having a single event which wasn't necessitated by antecedent causes, and that single event need not be an act on the part of some human agent. For the purposes of this paper, however, the kind of failure of the world to be fully deterministic at issue is just that kind in which the failure includes indeterministic human acts. So in what follows, I will speak of the world's failing to be fully deterministic and the world's containing indeterministic acts as if they were equivalent.

<sup>5</sup> The main impetus for the view that moral responsibility is compatible with causal determinism comes from a conviction that the world is characterized by causal determinism and a strong reluctance to give up belief in human free will or moral responsibility. I think that there is good reason, apart from considerations of moral responsibility, to question whether the world is characterized by causal determinism, and so I see no reason to reject the plausible view that moral responsibility and causal determinism are incompatible. For an excellent case against causal determinism, based on considerations in the philosophy of biology, see John Dupre, *The Disorder of Things: Metaphysical Foundations of the Disunity of Science* (Cambridge, MA: Harvard University Press, 1993).

(PAP) a person is morally responsible for doing an action *A* only if he had the ability to do otherwise than *A* (or if he had alternative possibilities to doing *A*).<sup>6</sup>

A standard strategy for trying to show that PAP is false is what has come to be known as a Frankfurt-style counterexample (FSC),<sup>7</sup> after Harry Frankfurt's well-known kind of counter-example against PAP. Some defenders of PAP have responded to the challenge of FSCs by arguing that, contrary to what FSCs are supposed to show, there remains a "flicker of freedom"<sup>8</sup> – that is, an alternative possibility for action – left to the agent in FSCs.<sup>9</sup> If that is so, then FSCs don't show that PAP is false, since the agent in a FSC is not an agent who lacks alternative possibilities.

I will argue that the flicker of freedom strategy is unsuccessful. In what follows, I will show that there is a dilemma for the proponents of the flicker of freedom strategy. Their strategy requires the supposition that doing an act-on-one's-own is itself an action of sorts. I argue that either this supposition is confused and leads to counter-intuitive results; or, if the supposition is acceptable, then it is possible to use it to construct a FSC in which there

---

<sup>6</sup> See, for example, Thomas Flint, "Compatibilism and the Argument from Unavoidability," *The Journal of Philosophy* 84 (1987), pp. 423–440. Flint makes clear how complicated the notion of having alternative possibilities for action is. For purposes of this paper, I will take an agent's having alternative possibilities for doing some action *A* as roughly equivalent to that agent's having the ability to do otherwise than *A*. Finally, PAP shouldn't be taken to cover cases of derivative responsibility, where an agent is causally determined to do an action, but where he himself is responsible for the state of affairs that causally determined him to do the action.

<sup>7</sup> See, for example, Harry Frankfurt, "Alternate Possibilities and Moral Responsibility," *The Journal of Philosophy* 66 (1969), pp. 829–839; reprinted in Harry Frankfurt, *The Importance of What We Care About* (Cambridge: Cambridge University Press, 1988), pp. 1–10.

<sup>8</sup> This is John Martin Fischer's phrase; see, for example, his *The Metaphysics of Free Will* (Oxford: Basil Blackwell, 1994), pp. 134–147.

<sup>9</sup> A different kind of defense against FSCs can be found in David Widerker, "Libertarian Freedom and the Avoidability of Decisions," *Faith and Philosophy* 12 (1995), pp. 113–118, and "Libertarianism and Frankfurt's Attack on the Principle of Alternative Possibilities," *The Philosophical Review* 104 (1995), pp. 247–261. I have argued against Widerker's position in "Libertarian Freedom and the Principle of Alternate Possibilities," Daniel Howard-Snyder and Jeff Jordan (eds.), *Faith, Freedom, and Rationality* (Rowman and Littlefield, 1996), and I recapitulate and develop that argument later in this paper. Van Inwagen himself has adopted a different defense against FSCs. He argues that if we construct PAP in a sufficiently careful way, FSCs don't show PAP in that formulation false. For present purposes, I am going to leave van Inwagen's argument to one side, because it relies on a somewhat unusual notion of what an agent can be taken to be responsible for. In a subsequent paper, I examine van Inwagen's view in detail and give a counter-example against it.

is no flicker of freedom at all. Either way, the flicker of freedom strategy is ineffective against FSCs.

Since the flicker of freedom strategy is the best defense of PAP I know, I conclude that FSCs are successful in showing that PAP is false. An agent can, therefore, act with moral responsibility without having alternative possibilities available to her.

If acting indeterministically did entail having the ability to do otherwise, then, of course, not having the ability to do otherwise would entail not acting indeterministically. And in that case anyone who rejected PAP would also be committed to rejecting the claim that moral responsibility entails indeterminism. But I will argue that it is perfectly possible to suppose that the agent in FSCs is acting indeterministically, even though he doesn't have the ability to act otherwise than he does. So even if moral responsibility entails indeterminism, indeterminism doesn't entail having alternative possibilities for action; and it is possible to reject PAP without denying that moral responsibility entails indeterminism.

So whereas van Inwagen presents his position as a package deal, in which the claim about moral responsibility's requirement of indeterminism rests on the claim about moral responsibility's requirement of alternative possibilities for action, I will argue that these two claims need not be connected. One can accept the former without the latter.<sup>10</sup> One can suppose that an agent who is morally responsible for some action is not causally determined to do what she does and yet that she does not have alternative possibilities with regard to that action.

Furthermore, supposing that we are at least sometimes morally responsible for our acts and that moral responsibility entails indeterminism seems to me sufficient for (at least one kind of) libertarianism.<sup>11</sup> If that is right, then what I say here can also be construed as an argument that libertarianism doesn't require a commitment to PAP. It also follows that the battle line between compatibilists and incompatibilists, which has concentrated on PAP, has been misdrawn.

Finally, at the end of the paper I consider reasons that might motivate libertarians to suppose that they need to be committed to some version of PAP, and I suggest that the basic intuitions behind libertarian commitment to PAP can in fact be satisfied better by requiring that a morally respon-

<sup>10</sup> One might also wonder whether an agent's having alternative possibilities for an act *A* is sufficient for her acting indeterministically and with moral responsibility. In a subsequent paper, I will argue that it is not sufficient.

<sup>11</sup> If it weren't for the tendency of some libertarians to take *free will* as equivalent to *the ability to do otherwise*, I would also be glad to phrase this claim this way: supposing that we at least sometimes act with free will and that acting with free will entails acting indeterministically seems to me sufficient for (at least one kind of) libertarianism.

sible agent<sup>12</sup> be the ultimate source of the acts for which he is morally responsible than that he have alternative possibilities for those acts.

#### FRANKFURT-STYLE COUNTEREXAMPLES

Frankfurt-style counterexamples can be constructed either for bodily actions, such as leaving a room, or for mental actions, such as an act of will to leave a room. In a FSC, a person *P* does an act *A* in circumstances that incline most people to conclude that *P* is morally responsible for doing *A*, but the circumstances involve some mechanism that would have operated to bring it about that *P* would have done *A* if *P* had not done *A* on his own. In the actual sequence of events presented in the counterexample, however, the mechanism does not operate, and *P* does do *A* on his own. So the counterexample is designed to make us think that *P* is morally responsible for doing *A* in the actual sequence of events although it is not the case that *P* could have done otherwise than *A*.<sup>13</sup>

Here is a FSC which is a revised version of one presented by John Martin Fischer.<sup>14</sup> Fischer's own version, like most FSCs, is vague about exactly how the coercive mechanism works and what it operates on. In one sense, this is perfectly acceptable, given the general purposes of FSCs; but for the sake of examining the "flicker of freedom" which is supposed to remain in a FSC it helps to spell out the details of the coercive mechanism a bit more and to consider the theory of mind that the FSC presupposes. For that reason, I've revised Fischer's example to make the operation of the coercive mechanism clearer. This revised version of Fischer's FSC features a neurosurgeon named "Grey," so call this FSC "(G)" after him.

(G) Suppose that a neurosurgeon Grey wants his patient Jones to vote for Republicans in the upcoming election. Grey has a neuroscope which lets him both observe and bring about neural firings which correlate with acts of will on Jones's part. Through his neuroscope, Grey ascertains that every

<sup>12</sup> Or an agent who acts with libertarian free will.

<sup>13</sup> See also my discussion of FSCs in "Intellect, Will, and Alternate Possibilities," reprinted in John Martin Fischer and Mark Ravizza (eds.), *Perspectives on Moral Responsibility* (Ithaca, NY: Cornell University Press, 1993), pp. 237–262, and "Persons: Identification and Freedom," *Philosophical Topics* 24 (1996), pp. 183–214.

<sup>14</sup> Fischer's version of this FSC can be found in his "Responsibility and Control," *The Journal of Philosophy* 89 (1982), p. 26; this paper is reprinted in John Martin Fischer (ed.), *Moral Responsibility* (Ithaca, NY: Cornell University Press, 1986), pp. 174–190.

time Jones wills<sup>15</sup> to vote for Republican candidates, that act of his will correlates with the completion of a sequence of neural firings in Jones's brain that always includes, near its beginning, the firing of neurons *a*, *b*, *c* (call this neural sequence "*R*").<sup>16</sup> On the other hand, Jones's willing to vote for Democratic candidates is correlated with the completion of a different neural sequence that always includes, near its beginning, the firings of neurons *x*, *y*, *z*, none of which is the same as those in neural sequence *R* (call this neural sequence "*D*"). For simplicity's sake, suppose that neither neural sequence *R* nor neural sequence *D* is also correlated with any further set of mental acts.<sup>17</sup> Again for simplicity's sake, suppose that Jones's only relevant options are an act of will to vote for Republicans or an act of will to vote for Democrats.

Then Grey can tune his neuroscope accordingly. Whenever the neuroscope detects the firing of *x*, *y*, and *z*, the initial neurons of neural sequence *D*, the neuroscope immediately disrupts the neural sequence, so that it isn't brought to completion. The neuroscope then activates the coercive neurological mechanism which fires the neurons of neural sequence *R*, thereby bringing it about that Jones wills to vote for Republicans. But if the neuroscope detects the firing of *a*, *b*, and *c*, the initial neurons in neural sequence *R*, which is correlated with the act of will to vote for Republicans, then the neuroscope does not interrupt that neural sequence. It doesn't activate the coercive neurological mechanism, and neural sequence *R* continues,

---

<sup>15</sup> By 'wills' in (G), I mean one mental act of willing, as distinct, say, from a persistent state of will of the sort Barkis had in mind when (in Dickens's *David Copperfield*) he told Davy to say "Barkis is willing."

<sup>16</sup> Not every philosopher of mind, of course, supposes that there are any regular correlations, even non-law-like correlations, between mental and neural states. If there are no regular correlations, or even if the regular correlations are violated a certain amount of the time, then Grey's neuroscope won't work. In that case, to make a suitable Frankfurt-style neuroscope, we would need to postulate a very fancy neuroscope or a much smarter neurosurgeon, so that the neurosurgeon can tell in every case what mental state will exist at the completion of any given neural process. I don't see, however, that the degree of fantasy in the neurobiology fantasy story affects its ability to serve as a Frankfurt-style counterexample. Furthermore, I also think that recent research in neurobiology strongly suggests there is a regular correlation between neural states and mental states for normal adult human beings, so that, for example, the mental state of seeing something regularly correlates with neural states in the occipital lobe; it doesn't correlate, not even a very small percent of the time, with states in, say, the cerebellum or the pituitary.

<sup>17</sup> The example doesn't require this simplifying assumption. If the firing of the initial neurons of each neural sequence were correlated with more than one set of mental acts, then the neurosurgeon's coercive mechanism would interfere with much more than it needs to interfere with for Grey's purposes, but this state of affairs doesn't alter the efficacy of the neuroscope in bringing about Grey's end.

culminating in Jones's willing to vote for Republicans, without Jones's being caused to will in this way by Grey.

And suppose that in (G) Grey does not act to bring about neural sequence *R*, but that Jones wills to vote for Republicans, without Grey's coercing him to do so.

It certainly seems as if Jones is morally responsible for his act of will to vote for Republicans, and yet it also seems true that it was not possible for Jones to do anything other than willing to vote for Republicans.

In this FSC, I am presupposing two things to which I want to call attention.

The first presupposition is that the mind is at least implemented in the brain and that therefore there is some correlation between mental states and neural states.

By saying that mental states are correlated with neural states, I mean to claim that there is a strong connection between a mental act or state and a neural state, but to leave general and vague the precise nature of that connection in order to make (G) and other FSCs like it compatible with a variety of different theories of relations between mind and brain. Those who think that the mental is identical to the physical can suppose that mental acts or states and sequences of neural firings are correlated because the mental acts or states *are* the neural states. Non-reductive materialists can take the correlation as some weaker relation, such as supervenience or emergence. Dualists willing to accept a correlation between mental and neural states might interpret the correlation as states of soul and body which are somehow connected. They might suppose, for example, that what happens in the soul is always mirrored at the same time by what happens in the body and vice versa, so that affecting the brain with drugs or other medical intervention is accompanied by a simultaneous alteration in the soul. (G) is therefore also compatible with some dualist theories of mind, namely, those that don't suppose mental acts are isolated in the soul, altogether unconnected in any way to sequences of neural firings.<sup>18</sup> (G)

---

<sup>18</sup> For a dualist theory compatible with supposing that the mind is implemented in the brain, see the position I ascribe to St. Thomas Aquinas in "Non-Cartesian Dualism and Materialism without Reductionism," *Faith and Philosophy* 12 (1995), pp. 505–531. Furthermore, the agents at issue in this paper are embodied human beings in this life, but nothing in the presuppositions of my FSC precludes a dualist from holding that there is an afterlife in which human beings are disembodied altogether and yet have mental acts and states. Finally, it may be the case that agent-causation is a sort of dualism immune to this FSC; but if it is, it is so because it shares with Cartesian dualism at least this claim, namely, that there can be a change in a state of soul without there being a simultaneous change in the brain (in this connection, see also footnote 19).

does assume that, whatever else may be true about the connection between the mental and the neural, if the neural sequence correlated with a mental act or state exists, the mental act or state does too. But this is an assumption which should be acceptable to dualists of the sort just described (and, even more obviously, to those theorists who take the mental to be identical with, supervenient on, or emergent from the neural).

Those who think there are no psycho-physical laws can also be accommodated here. The correlation in question need not be law-like. All Grey's neuroscope needs in order to operate is a current correlation in Jones between the acts of will to vote in certain ways and the neural states at issue. But this correlation need not hold across human beings; it need not even hold throughout Jones's lifetime, as long as it characterizes him in the period in which Grey is investigating and manipulating him.

The neurological fantasy story in (*G*) is therefore compatible with most current theories of mind, provided that they tie mental states to neural states in some suitable way.<sup>19</sup>

The second presupposition is that the correlation between a mental act or state and the firings of neurons is a one-many relation. When I suddenly recognize my daughter's face across a crowded room, that one mental act of recognition, which feels sudden, or even instantaneous, is correlated with many neural firings as information from the retina is sent through the optic nerve, relayed through the lateral geniculate nucleus of the thalamus, processed in various parts of the occipital cortex, which take account of figure, motion, orientation in space, and color, and then processed further in cortical association areas. Only when the whole sequence of neural firings is complete, do I have the mental act of recognizing my daughter. Whatever neural firings are associated with an act of will, I take it that in this case, as in all others, the correlation between the mental act and the firing of the relevant neurons is a one-many relation.<sup>20</sup>

<sup>19</sup> It must be conceded, of course, that a FSC of this sort, which relies on intervention in the brain, couldn't be constructed for agents who are disembodied souls, so this FSC isn't compatible with Cartesian dualism. It doesn't follow, however, that it isn't possible to construct FSCs which are compatible with Cartesian dualism. Although the sort of FSCs currently in fashion incorporates science fiction stories involving neuroscopes, I see no reason why FSCs couldn't be built around theology fiction stories instead, in which divine omnipotence and omniscience accomplish what the neurologist's skill and tools do in more common FSCs. In that case, with some assumptions about the nature of omniscience, we could build FSCs even for disembodied souls.

<sup>20</sup> What kind of one-many relation this is depends on what theory of the relationship of mind to brain one adopts. For those who think that mental states are identical to neural states, for example, the correlation between mind and brain has the implication that a mental act is temporally extended throughout the microseconds it takes for all the neurons in the correlated sequence to fire. (Something needs to be said to explain why, even on this



It is important to be clear about this point. If the firing of the whole neural sequence correlated with a mental act is not completed, the result isn't some truncated or incomplete mental act. It's no mental act at all. If the neural sequence correlated with my recognizing my daughter's face across a crowded room is interrupted at the level of the thalamus, say, then I will have no mental act having to do with seeing her. I won't, for example, think to myself, "For a moment there, I thought I saw my daughter, but now I'm not sure." I won't have a sensation of almost but not quite seeing her. I won't have a premonition that I was about to see her, and then I mysteriously just don't see her. I will simply have no mental act regarding recognition of her at all. To suppose that there could be some sort of mental act, truncated, incomplete, or otherwise defective, when there is no completed neural sequence correlated with that mental act, is to accept some version of Cartesian dualism. It is to suppose that there can be a mental act without there being a completed neural sequence correlated with that mental act. So if the neural sequence correlated with a mental act is interrupted, then that mental act doesn't occur. If there is any mental act at all in those circumstances, it will occur only because there is some other completed sequence of neural firings correlated with *that* mental act. So although a mental act such as a decision may feel, subjectively, as if it is simple and instantaneous, the neural sequence with which it is correlated is neither.

These presuppositions together have the result that there is no act of will in an agent unless and until the correlated sequence of neural firings in that agent's brain is completed.<sup>21</sup>

---

view, there is no mental act without a completed neural sequence. Perhaps the proponents of this view might want to say that a mental act has subjectively indiscernible components and that the act correlated with the completed neural sequence doesn't exist unless and until all its components exist.) On other theories of the relation of mind to brain, it could be the case that the mental act comes into existence simultaneously with the firing of the last neuron in the sequence.

<sup>21</sup> On my presuppositions, then, if a neural sequence and a mental act or state are correlated, the neural sequence exists if and only if the correlated mental act or state exists. To ward off the sort of confusion which sometimes arises in this connection in philosophy of mind, it is probably helpful to add that nothing in my presuppositions makes it necessary that mental acts and states be correlated with neural sequences – there might be creatures for whom the mental is correlated with states of silicon instead – or that there be one and only one neural sequence which is correlated in a law-like way, for all human beings or even within the life of just one human person, with one particular mental act or state. All (G) requires is that a particular embodied human being in this world be such that he has some mental state or engages in some mental act if and only if the neural sequence correlated with that act or state in him is completed. And, as I have explained, this is a position which even some dualists can accept. If a mental act or state in the soul is simultaneously accompanied by a neural state in the brain and vice versa, then if that neural

## INDETERMINISM AND FRANKFURT-STYLE COUNTEREXAMPLES

It is my contention that, as I have formulated this FSC, there is nothing in it which requires us to deny that the potential victim of neurological coercion acts indeterministically.<sup>22</sup> When Jones wills to vote for Republicans, Grey is not intervening to cause any mental state or neural firing in Jones's brain. And the assumption that *nothing* acted causally on Jones to produce the act in question, that Jones acted indeterministically, is not incompatible with anything specifically postulated in or required for (G).<sup>23</sup>

David Widerker has put forward an argument which might be thought to undermine this last claim. He maintains that all FSCs presuppose that *something* causes the victim to act as he does even when the coercive mechanism doesn't operate.<sup>24</sup> That is because the neuroscope has to respond to something in the victim; but, Widerker argues, what it responds to must be causally necessary for the desired act on the victim's part. If this were not the case, Widerker maintains, the victim would after all have a possibility for doing otherwise. So, for example, in some versions of FSCs, the counterfactual intervener detects an inclination to an act *A* on the victim's part, and in consequence the intervener intervenes to produce not-*A* in his victim. If, however, the victim could suddenly do *A* without having previously had an inclination to do *A*, then the intervener's mechanism couldn't operate in time; and in that case, contrary to what the FSC is supposed to show, the victim would have had it in his power to do otherwise. So the inclination must be a causally necessary condition

---

state exists, it is true that the correlated mental act or state does also, and if the neural state does not exist, it's true that the correlated mental act or state doesn't exist either.

<sup>22</sup> This is not to say, of course, that anything in the example requires that Jones be acting indeterministically. My contention is only that nothing in the example rules it out. Where in the example indeterminism might come in depends on one's view of indeterminism. But nothing in my example precludes supposing, for example, that the neural sequence correlated with the mental act in question simply begins to fire, as if by magic, without being caused by anything at all. I myself don't suppose that the rejection of causal determinism requires any introduction of magic of this sort, but I call attention to this possibility to underscore the point I want to make in this section. Nothing in the example itself precludes even this magical sort of indeterminism. *A fortiori*, nothing in the example itself requires the assumption of causal determinism.

<sup>23</sup> Someone might suppose that if neural firings are correlated with mental acts, then it must be the case that the neural state, and consequently also the mental act, is caused by something outside the agent. That is, some people might suppose that what happens in the brain can be understood only as part of a fully deterministic world. I have argued at length against such a claim in "Libertarian Freedom and the Principle of Alternate Possibilities," *op. cit.*

<sup>24</sup> Widerker, "Libertarian Freedom and the Avoidability of Decisions" and "Libertarianism and Frankfurt's Attack on the Principle of Alternative Possibilities."

for the victim's doing *A*. But in that case not having an inclination to do *A* is causally sufficient for an agent's doing not-*A*. And so, contrary to appearances, FSCs presuppose a causally determined agent.

I agree with Widerker that if the coercive mechanism in a FSC is responsive to something which is both a mental state and antecedent to the desired act on the victim's part, then a case can be made that the FSC is not compatible with supposing that the victim acts indeterministically in the actual sequence in which the coercive mechanism doesn't operate. But, as (G) makes clear, for contemporary theories of mind other than extreme Cartesian dualism, it is possible that the coercive mechanism is responsive not to a *mental state antecedent* to the desired act of will, but rather to a *neural state* which is correlated with *that very act of will*. In FSCs of this sort, nothing about the nature of the coercive mechanism requires that, in the actual sequence where that mechanism doesn't operate, the victim's act be causally determined by a mental act antecedent to that act of will. Even if we supposed that Jones's act of will in the actual sequence is caused by nothing at all, Grey's neuroscope could still work, since it is responsive only to the neural firings which are correlated with Jones's very act of will. Consequently, it is possible to build FSCs in such a way that they evade Widerker's criticism.<sup>25</sup>

A very similar point to Widerker's is made by Robert Kane, who argues that an indeterministic choice on an agent's part can't be, as he says, "Frankfurt controlled," because "if the controller does not intervene, the choice remains undetermined up to the moment that it occurs."<sup>26</sup> Kane

<sup>25</sup> In correspondence, William Hasker remarks about my argument against Widerker: "Given the initial sequence '*R*' together with other relevant conditions that obtain, is the completion of the 'vote-Republican act of will' nomologically necessary, or not? If not, then it is still possible, even *after* the sequence '*R*' occurs, for the agent to come to her senses and vote Democratic anyway. . . . in this case the agent does have an alternative possibility. But if the completion of the act of will *is* nomologically necessary under those circumstances, what we have to conclude is that *the effective decision has already been made before 'R' occurs*. So now, if the decision is to be indeterministic . . . , the indeterminism must enter the process *before 'R' occurs*." These remarks are helpful because they show how hard it is really to hold in the forefront of one's mind the notion that the mental and the neural are correlated. In my example, the neural sequence *R* is correlated precisely with the "effective decision" to vote Republican; but if something in my example implied or suggested that the effective decision was made before *R* occurs, then *that* effective decision would itself be correlated with a neural sequence, and my remarks about *R* would apply to that neural sequence. The alternative is to suppose that decisions occur in the soul, and that there need be no change of brain states when those decisions occur. Although some libertarians hold such a version of Cartesian dualism, it is not generally supposed that Cartesian dualism is a prerequisite for libertarianism.

<sup>26</sup> Robert Kane, *The Significance of Free Will* (Oxford: Oxford University Press, 1998), p. 192.

says, “[In a FSC] the controller, Black, plans to make Jones do *A*. But he waits to see if Jones is going to do *A* on his own and only intervenes if Jones is about to do *B* instead. But if *A* is a SFW [self-forming willing] (say, a choice), the controller faces a dilemma in carrying out such a strategy. Since the SFW is preceded by an indeterminate effort, it is undetermined whether choice *A* or *B* will occur until one or the other of them actually does occur. The controller cannot know which one is going to occur beforehand unless he predetermines one of them to occur. He can therefore wait until he finds out whether the agent will do *A* or *B*, but then it is too late to control the choice. . . . If they [Frankfurt-style controllers] tamper with our brains, shutting off the indeterminacy prior to choice, they can control our choices, but not if the efforts remain indeterminate and the choices undetermined up to the moment they occur.”<sup>27</sup>

Kane here makes explicit what Widerker also assumes, namely, that a mental act such as an act of willing is something indivisible which occurs at a moment. I agree that subjectively considered, from the point of view of the one willing, a mental act of willing seems to occur at once and to be indivisible in its nature. But if the mind is implemented in the brain, then any mental act, even an act of willing, is correlated with a neural state which is not indivisible and which occurs over a period of time, however small. My act of recognizing my daughter across a crowded room also seems to me to be something which is indivisible and which occurs at once. But it is correlated with a neural state which is neither. If a crazy neurosurgeon with futuristic technology wanted to keep me from recognizing my daughter, his neuroscope would not have to respond to any mental state antecedent to my recognition of my daughter or even to any antecedent state of the world which might be thought to cause my recognition of my daughter. The neuroscope can be responsive only to the sequence of neural firings which is correlated with the very recognition itself. If it aborts that sequence before the sequence is completed, then the neuroscope will prevent my having the mental act of recognition. But it will do so without responding to or acting on anything antecedent to the neural sequence which is correlated with the act itself. Even if we stipulated that – somehow or other – that act arose without being caused to arise by anything whatever, if in some mysterious way the neural sequence correlated with the act just began on its own, the neuroscope would still have plenty of room to operate.

If we assume that acts of will occur only in a soul which is, at least as regards willing, disconnected from the brain, then the Widerker/Kane point will hold. But if we accept any more plausible theory of the relation

---

<sup>27</sup> Kane, *The Significance of Free Will*, pp. 142–143.

between mind and brain, the claim that “the choice remains undetermined up to the moment that it occurs” does not entail that the choice cannot be “Frankfurt controlled.”<sup>28</sup>

Since, as far as I know, no one has given a plausible reason other than that given by Widerker and Kane for supposing that FSCs require an agent who is causally determined even in the actual sequence, I conclude that, unless we are prepared to accept extreme Cartesian dualism, FSCs such as (G) – Frankfurt stories in which the Frankfurt controller interrupts neural sequences rather than preempts his victim’s act of will – are compatible with supposing that the victim acts indeterministically in the actual sequence. Therefore, if FSCs really do rule out alternative possibilities for the victim, they also show (to everyone except extreme Cartesian dualists<sup>29</sup>) that an agent who lacks alternative possibilities with regard to an action need not be causally determined to do that action. Consequently, if it’s true that the victim in FSCs has no alternative possibility to doing what he does, then FSCs demonstrate that acting indeterministically doesn’t entail having alternative possibilities for action.

#### A DEFENSE OF PAP: THE FLICKER OF FREEDOM

It remains an important question, however, whether the victims in FSCs really lack alternative possibilities for action. Not everyone agrees that they do. Characterizing the views of these defenders of PAP, Fischer says,

upon closer inspection it can be seen that, although they [the FSCs] do not involve alternative possibilities of the normal kind, they nevertheless may involve *some* alternative possibilities. That is to say, although the counterfactual interveners eliminate most alternative possibilities, arguably they do not eliminate *all* such possibilities: even in the Frankfurt-style cases, there seems to be a “flicker of freedom.”<sup>30</sup>

<sup>28</sup> Kane argues that since indeterministic choices cannot be “Frankfurt controlled,” it is not possible that there be an agent who makes indeterministic choices but who never has the ability to do otherwise, who is under the control of a counterfactual intervener throughout his whole life (Kane, *The Significance of Free Will*, p. 143). But if I am right in my argument that even an indeterministic choice can be “Frankfurt controlled,” then it is also possible that there be an intervener who monitors all the choices of his victim, from birth to death, but who intervenes in none of them because all the victim’s choices are the ones the counterfactual intervener approves of. It would be easy enough to construct a theology fiction story in which God takes the place of the neurologist in the more customary science fiction stories and in which God “Frankfurt controls” all of Jones’s choices without causing any of them. For these reasons, I think that Kane’s argument for retaining a limited version of PAP is unsuccessful.

<sup>29</sup> I will stop inserting this qualification of the claim being made here; it should be understood in what follows.

<sup>30</sup> Fischer, *The Metaphysics of Free Will*, p. 134.

The intuition that there is after all an alternative possibility which remains for the victim in FSCs can be based on either of two features of FSCs.

The first feature is that the intervener's mechanism needs to be responsive to something in the victim. Whatever it is that the neuroscope responds to, it seems to some philosophers that the victim has the power to do or not to do *that*. As FSCs have generally been formulated, the neuroscope is responsive to some mental state which is antecedent to the desired act on the victim's part. It can appear that the victim has the ability to form or not to form that antecedent mental state and therefore that there are alternative possibilities after all for the victim.

I am inclined to agree with Widerker, however, that such FSCs covertly suppose that the victim is causally determined to do as he does even in the actual sequence in which the coercive mechanism doesn't operate. Consequently, such FSCs beg the question which has been at issue between the opponents and the defenders of PAP, namely, whether an agent who is causally determined to do what he does can be morally responsible for it. It therefore is much less interesting to determine whether or not the victim in such FSCs actually has alternative possibilities open to him. For this reason, I will not consider any further defenses of PAP which base their argument on the idea that the victim in FSCs does have alternative possibilities for action on this first feature of FSCs.

The second feature is that in the actual sequence the victim does some act *A* on his own, as it were, and in the alternative sequence he is caused to do *A* by the counterfactual intervener. Consequently, it seems that there is an alternative possibility open to the agent after all. He can do *A* on his own, or he can fail to do *A* on his own. Defenses of PAP based on this feature of FSCs are not so easily dismissed.

Fischer grants that there are alternative possibilities of this sort for the victim in FSCs, but he argues that they aren't enough to ground moral responsibility. He points out that in the alternative sequence when the intervener manipulates the victim, the victim himself doesn't engage in any act at all, let alone an act which manifests the sort of control over his action which is supposed to be present in morally responsible acts.<sup>31</sup> Therefore, PAP isn't rescued from FSCs by noticing that the victim in FSCs has alternative possibilities of this sort. Fischer is thus supposing that defenders of PAP take moral responsibility as requiring not only an alternative *action* on the agent's part, but even an action for which the agent is morally responsible. Since it is clear that there is no such action on the victim's part in the alternative sequence of FSCs, FSCs are still successful, Fischer thinks, in showing that PAP is false.

---

<sup>31</sup> Fischer, "Responsibility and Control," p. 31.

It is, however, open to the defenders of PAP to maintain that moral responsibility doesn't require the sort of *robust* alternative possibilities Fischer supposes, in which the alternative possibilities available to the agent are both acts and acts for which the agent is responsible. As long as there are any alternative possibilities at all available to the agent, they might argue, the requirement for PAP is met.<sup>32</sup> For my purposes here, it isn't necessary that Fischer's view of what is required for the defense of PAP be right, although in fact I think it is.<sup>33</sup> I want to approach the issue of the flicker of freedom somewhat differently. I want to look more closely at the nature of the alternative possibility that is supposed to remain in the flicker of freedom.

#### DOING AN ACT ON ONE'S OWN

Flicker of freedom arguments of the sort at issue here are based on the distinction between the victim's doing something on his own and his doing it as a result of the operation of a coercive mechanism. There certainly is such a distinction; and, as I explain in the last section of this paper, I agree with the proponents of the flicker of freedom strategy that it is a distinction which makes a great difference to assessments of moral responsibility. But the question here is whether this distinction saves PAP. Proponents of this flicker of freedom strategy, who think it does save PAP, are apparently thinking along these lines. In FSCs, the victim performs an action in the actual sequence which he fails to perform in the alternative sequence, namely, doing some act on his own. Now an agent's failure to perform an action is clearly different from the agent's performing that act. What reason could there be for thinking that this alternative is not a real alternative possibility for the victim?<sup>34</sup>

<sup>32</sup> In correspondence, William Rowe raised an objection to Fischer along these lines.

<sup>33</sup> In footnote 35, I argue for the claim that PAP must be interpreted as requiring alternative possibilities each of which is itself an action, although I am happy to count certain sorts of omissions as actions, too.

<sup>34</sup> In correspondence, William Rowe dissociates his own flicker of freedom argument from this strategy. He says, "The difference [between the actual and the alternative sequence in a FSC] is that in the actual sequence he [the victim] is the cause of that act of will and resulting action (having the power not to be the cause), whereas in the alternative sequence the intervener is the cause of the agent's willing and acting as he does. . . . An agent's causing his volition isn't itself a further action." On Rowe's approach, a person's causing his volition isn't itself an action on that person's part; it isn't something the victim in FSCs does. In that case, however, what the victim does, the victim's actions, in the actual sequence and the alternative sequence are just the same. What differs is only the mode of the action, which is caused by the agent in the actual sequence but not in the alternative sequence. I agree with Rowe that the difference in the mode of the action in the actual and

Designate as “*W*” the act of the victim which in FSCs the counterfactual intervener desires, such as the act of willing to vote for Republicans in (*G*). As the proponents of the flicker of freedom defense point out, quite correctly, the victim in FSCs does *W* on his own in the actual sequence but not in the alternative sequence. It is important to see, however, that in order to turn this feature of FSCs into a defense of PAP, the flicker of freedom proponents must make two assumptions. They must take the victim’s doing *W*-on-his-own as something which the victim *does*, and they must suppose that doing *W*-on-his-own is not identical to doing *W*. Both these assumptions are necessary to their case. If doing *W*-on-his-own weren’t an *action* the victim does, then there wouldn’t be something the agent does in the actual sequence but omits to do in the alternative sequence, as the flicker of freedom proponents argue.<sup>35</sup> And if doing *W*-on-his-own weren’t different from doing *W*, then what the victim does in the actual and the alternative sequence would be identical, and the victim wouldn’t have *alternative* possibilities available to him.

---

the alternative sequences is both significant and enough to ground moral responsibility, as I explain in the last section of the paper, where I discuss this sort of construal of FSCs. But it is hard to see that this is a defense of PAP against FSCs, since on this flicker of freedom argument, the agent is not able to act otherwise than he does. As Rowe explains it, *what* the victim does is the same in the actual and the alternative sequences. But, then, the victim isn’t in fact able to act otherwise than he does. Construing the difference between the acts of will in the actual and alternative sequences as Rowe does, as a matter of the mode of action, constitutes a reason for thinking that an agent must be the ultimate source of his action in order to be morally responsible for what he does, as I explain in the last section of this paper; but it doesn’t give any reason for thinking that when an agent is the ultimate source of his action, he must also have the ability to act otherwise than he does.

<sup>35</sup> In fact, I think that it is also necessary for their case that the omission count as an action. To see that this is right, consider (*G*) again. There is certainly a possible world in which Grey gets distracted at the last minute by some great personal crisis, so that he no longer cares which way Jones votes. And suppose that in that world Jones himself also has a great personal crisis, so that the election just goes right out of his head, too. In that world, Jones doesn’t do *W*-on-his-own. He doesn’t do *anything* as regards voting. That possible world is accessible to the actual world as it is described in (*G*), so that Jones’s not doing anything at all about the election is a real possibility for Jones. Consequently, there is another alternative sequence for Jones in (*G*), namely, the sequence in which Jones forgets all about the election, doesn’t vote for anybody, and therefore doesn’t vote for Republicans on his own. Here Jones omits to do *W*-on-his-own, too; in fact, he even omits to do *W* itself. But surely no one would suppose that the existence of this possibility shows that in (*G*) Jones has alternative possibilities available to him. Why is that? I think the answer is that, although there are some sorts of omissions which can be considered actions, too, the sort of omission being described here isn’t one of those. PAP, however, is a principle about alternative possibilities for action. So if the alternative possibility for the victim in a FSC is an omission which can’t be construed as an action in any sense, then pointing to the existence of that alternative possibility can’t constitute a defence of PAP. For these reasons, among others, I think Fischer is right in his interpretation of PAP.



Suppose we call any action of this kind – doing *W*-on-his-own – “*O*.” Then the preceding discussion makes it clear that, for proponents of this flicker of freedom defense, there are two things that must be true of any *O*. Any *O* must be an action, and it must be a different action from *W*.

Now it isn’t clear that it makes sense to take *O* as an action at all, and there are counterintuitive consequences of doing so. To see this, consider, for example, an analogue to (*G*) in which the scenarios in the actual and the alternative sequences are switched:

Analogue (*A*): In the actual sequence, Grey activates his coercive neurological mechanism and himself brings it about that Jones wills to vote for Republicans. In the alternative sequence, Grey monitors Jones but doesn’t activate the coercive neurological mechanism, and Jones wills to vote for Republicans on his own, without any coercion on Grey’s part. (We can add to this example, if we like, the stipulation that the worlds of the actual sequence and the alternative sequence are not fully deterministic and that Jones typically acts indeterministically in either one.)

If there were two alternative possibilities available to the victim in a standard FSC such as (*G*), as the proponents of the flicker of freedom defense maintain, then there ought to be the same two alternative possibilities available to Jones in (*A*). It’s just that the possibility for action available to Jones in the actual sequence in (*G*) – – *O* – – is now the possibility available to him in the alternative sequence in (*A*), and the possibility available in the alternative sequence in (*G*) – – *W* – – is the possibility available in the actual sequence in (*A*).

That is, if *O* is an action in its own right, then we should be able to schematize Jones’s alternative possibilities in (*G*) and in the analogue (*A*) this way:

(G)		(A)	
<u>Actual seq.</u>	<u>Alt. seq.</u>	<u>Actual seq.</u>	<u>Alt. seq.</u>
<i>O</i>	<i>W</i>	<i>W</i>	<i>O</i>

But it seems clearly false that in (*A*) Jones has alternative possibilities for action or the ability to do otherwise than he does. In the actual sequence in (*A*) Jones would be entirely within his rights in claiming, afterwards, that he couldn’t have done otherwise than he did, and he wouldn’t be moved to rescind that claim by our insistence that there was an alternative possibility for his action in the alternative sequence in which he does *O*. Furthermore, this is so even though in (*A*), unlike (*G*), the possibility in the alternative sequence includes an action on Jones’s part for which he clearly *is* morally responsible, because doing *O* is doing *W* on his own, without any interference from Grey or anybody else.

Someone might object here that, for one reason or another, the relevant alternative to *O* ought to be not *W*, but a simple failure to perform *O*. I don't think this objection is right, but for the sake of argument, suppose that it is. Then everything said here about *O* and *W* will apply to *O* and the failure to do *O*. In that case, we can schematize (G) and (A) this way:

(G)		(A)	
<u>Actual seq.</u>	<u>Alt. seq.</u>	<u>Actual seq.</u>	<u>Alt. seq.</u>
<i>O</i>	–	–	<i>O</i>

Clearly, even on the objection, it remains the case that there are alternative possibilities for Jones's action in (A) if it is true that there are such alternative possibilities in (G). In fact, as I pointed out just above, the possibility in the alternative sequence of (A) is a possibility for an action over which, according to the flicker of freedom proponents, Jones has the right sort of control and for which he is morally responsible. To suppose that Jones has alternative possibilities for action in (A), however, is clearly a counter-intuitive result, and it stems from taking doing *W*-on-one's-own as an action.

So although the proponents of the flicker of freedom defense of PAP need to take doing an act *W* on one's own as an act in its own right, consideration of cases such as (A) shows that there is something wrong about doing so, and the counter-intuitive results arise from considering doing *W*-on-his-own as an action that the victim in FSCs does.

#### FRANKFURT-STYLE COUNTEREXAMPLES WITHOUT A FLICKER OF FREEDOM

For those who are unpersuaded that there is anything mistaken about taking doing *W*-on-one's-own as an action, however, there is a different way to undermine the flicker of freedom strategy. Suppose that, for the sake of argument, we accept that doing *W*-on-one's-own is itself an action. In that case, we can construct a FSC in which there isn't even a flicker of freedom.

If *O* is an act on the victim's part, as we are assuming here for the sake of argument, then *O* must either be or be caused by a mental act. If *O* were neither a mental act nor caused by a mental act on the victim's part, it's hard to see how *O* could count as something that the *agent* does. For simplicity's sake, suppose that *O* is a mental act. Now, unless we accept Cartesian dualism, we will have to grant that *O* is correlated with a neural sequence in the brain; designate that neural sequence "*S*." There is nothing about the firing of the neurons in any neural sequence in the brain of any

human person such as Jones which makes it in principle impossible for someone or something external to Jones to bring the firing of that sequence about. Therefore, the firing of *S*, too, can be produced by an intervener. Consequently, *O*, which is correlated with *S*, can be brought about by an intervener just as *W* can. In that case, we can build a FSC in which the presence or absence of *O* is under the control of an intervener, just as *W* is in (*G*). In such a FSC, it is clear, there will not be for the victim an alternative possibility to doing *O*; the victim does *O* in both the actual and the alternative sequence.

To show that even in such a FSC there is an alternative possibility for action would require showing that doing *O*-on-his-own is an action and different from doing *O*. but doing *O*-on-his-own is just doing *W*-on-his-own-on-his-own, and it's hard to know what such an act is supposed to be. Even if we could give some coherent description of doing *O*-on-his-own, however, the same argument as before would apply. Doing *O*-on-his-own would be an action different from doing *O*, and we could construct a FSC in which the agent has no alternative but to do *O*-on-his-own.

Furthermore, we can construct one FSC in which the counterfactual intervener desires not just some act *W* on the part of the victim but also the further act *O*, as well as the act of doing *O*-on-his-own if there is such an act, and any further iterated acts of doing on one's own. We can stipulate that the counterfactual intervener controls all these acts in virtue of controlling the firings of neurons in the neural sequences correlated with each of these acts. If the victim doesn't do these acts, the coercive neurological mechanism will produce them. In such a FSC, there are no alternative possibilities for action of any sort on the part of the victim.

So, even if, for the sake of argument, we accept the notion of doing *W*-on-one's-own as an action, there is a serious problem for the proponents of the flicker of freedom defense. If doing *W*-on-his-own *is* an action which is different from *W*, then *that act* – – *O* – – is something that we can build a FSC around; and in the newly constructed FSC there will be no alternative possibility to doing *O* for the victim.

#### AN OBJECTION

In response to the preceding argument, someone might object that the FSC in which the neurologist causes the victim to do *O* in the alternative sequence is incoherent. "*O*" is introduced as an abbreviation for "*W*-on-one's-own," and the abbreviation, the objector might argue, masks a fundamental confusion in the FSC. According to the FSC, Jones does *O* in the actual sequence; but if he hadn't done *O*, Grey would have brought it about that Jones did *O*. Since Jones's doing *O*, however, is Jones's doing

*W-on-his-own*, to suppose that Grey can bring it about that Jones does *O* is to suppose that Grey can bring about Jones's doing *W-on-his-own*. In that case, the act Grey brings about is *W done by Jones on his own and caused by Grey to do so*. But this is an incoherent description. No act can be both done by an individual on his own, that is, without his being caused to do so by an intervener, and also done by that individual as a result of his being caused to do so by an intervener. So the FSC in which Grey brings it about that Jones does *O* is incoherent, and its incoherence is only thinly veiled by its use of "*O*" as an abbreviation for "*W-on-one's-own*."

Here I want to grant the objector's claim that the description of the act in question is incoherent, but I want to deny what the objector supposes is an implication of this claim, namely, that the FSC in which Grey brings it about that Jones does *O* is incoherent.

Remember that the FSC in question takes as an assumption that doing *W-on-one's-own* is an action, and suppose, for the sake of simplicity, that doing *W-on-one's-own* is a mental action. If we are Cartesian dualists, then we will suppose that this mental action takes place in a soul and only in a soul. The soul might subsequently have some effects on the brain, but the mental action itself occurs just in the soul, disconnected from the brain. But if we hold any theory of the mind friendlier to contemporary neurobiology than Cartesian dualism, we will have to suppose that any mental action is correlated with a neural sequence in the brain, in the way described above. If the mental and the neural are correlated in this way, then Jones's mental act of doing *W-on-his-own* also has a neural sequence correlated with it.

There is, however, no reason why any neural sequence whatsoever can't be fired by an external intervener such as Grey. Therefore, Grey can also bring about the firing of the neural sequence correlated with Jones's mental act of doing *W-on-his-own*. If Grey does so, then Grey will also bring about the mental act we have described as Jones's doing *W-on-his-own*. What follows from the recognition that the description of the act as *Jones's doing W on his own and caused to do so by Grey* is incoherent is only that the act is misdescribed in this way, not that Grey can't bring about the act.

To see this point, consider the A. D. White house at Cornell. It's quite correctly described as *the building which is across the street from Goldwin Smith Hall*. The description *the building which is across the street from Goldwin Smith Hall and which is not across the street from Goldwin Smith Hall* is clearly incoherent. And so one can imagine someone wondering whether the fact that it is correct to describe the A. D. White house as *the building which is across the street from Goldwin Smith Hall* means that nothing can happen to Goldwin Smith Hall while the A. D. White house stands. But, of course, this is absurd. The very same house, which

is now correctly described as across the street from Goldwin Smith Hall, would remain in existence even if Goldwin Smith Hall were to blow up. If Goldwin Smith Hall blew up, it would not be true that the same house now had an incoherent description, or that we had a different house from the one we had before, or that the house hadn't been correctly described before as *the house which is across the street from Goldwin Smith Hall*. All that would be true is that the description under which we picked out the house before, while Goldwin Smith was still standing, is not a description under which we can pick out the house once Goldwin Smith is blown up.

In the same way, the very mental act which we can pick out, quite correctly, under the description *Jones's doing W-on-his-own*, is an act which Grey can cause Jones to do. But, of course, if Grey brings it about that Jones does that act, we can no longer pick out the act under the description *Jones's doing W-on-his-own*. It was correctly described that way when Grey didn't cause Jones to do it; it can't be picked out under that description when Grey brings it about that Jones does it. But nothing in this fact means that Grey can't cause Jones to do this act, any more than the fact that the A. D. White house is correctly described as *the house which is across the street from Goldwin Smith* and cannot be coherently described as *the house which is across the street from Goldwin Smith and not across the street from Goldwin Smith* means that Goldwin Smith can't blow up as long as the A. D. White house stands or that there would be a different house if Goldwin Smith blew up.

So if Jones's doing *W-on-his-own* is a mental act, then, like any other mental act, it is correlated with a neural sequence. Grey, who can bring about neural sequences, can bring about this very neural sequence; and if he does so, then he brings about this very mental act which we otherwise correctly describe as Jones's doing *W-on-his-own*. If Grey brings it about that Jones does this act, we can no longer pick out the act under that description. But what changes when Grey causes Jones to act is the description of the act, not the act itself.

The alternative is to suppose that mental and neural states are dissociated, so that Jones could be in the very same neural state he is when he does *W-on-his-own* but not be engaged in the same mental act or that Jones could do *W-on-his-own* without there being any correlated sequence of neural firings.<sup>36</sup> To suppose either of these things, however,

---

<sup>36</sup> Here, as elsewhere in this paper, I am assuming that what is at issue is embodied human beings in this world. A caveat of this sort is needed because there are philosophers who take it to be logically possible that a person be in a physical state identical to the state Jones is in when he has some mental state and yet have no mental states of any sort at all. But what is at issue in my example is just human beings as they are in this world; and

would require adopting a Cartesian dualism in which mind and brain are disconnected.<sup>37</sup>

So although the description of an act as *Jones's doing W on his own and caused to do so by Grey* is incoherent, there is nothing incoherent about supposing that Grey causes the act which can be correctly picked out under the description *Jones's doing W-on-his-own* when Grey doesn't bring that act about. Consequently, nothing in the description of the act Jones does as *doing W-on-his-own* shows that the FSC in which Grey causes Jones to do *O* in the alternative sequence is incoherent.

For these reasons, I think that this objection fails.<sup>38</sup>

### SOME CONCLUSIONS

Consequently, either the notion that doing *W-on-one's-own* is itself an action – on which the flicker of freedom defense rests – is mistaken, or we can construct FSCs in which there is no alternative possibility even for actions such as doing *W-on-one's-own*. Either way, the proponents of the flicker of freedom strategy are not successful in showing that there are in fact alternative possibilities for the victim in FSCs.

It seems to me therefore that Frankfurt was right in the first place: FSCs show that PAP is false.

Of course, to say that PAP is false is not to say that, outside science fiction stories, there *are* morally responsible agents who never have *any* alternative possibilities for action. Rather, what I think the preceding arguments show is only that having alternative possibilities for action is not essential to moral responsibility. It's compatible with this conclusion that having alternative possibilities is what medieval logicians would have called an associated accident of moral responsibility: something which in our world generally accompanies being morally responsible even if it isn't required for it.

I have also argued that there is nothing in FSCs as I have constructed them here which precludes supposing that the victim is acting indeter-

---

even a dualist such as David Chalmers, who argues for the logical possibility of zombies, maintains that in this world, mental and neural states are correlated in the way required for my point here. See David Chalmers, *The Conscious Mind* (Oxford: Oxford University Press, 1996), esp. Chapter 3. See also the point about FSCs and disembodied souls in footnote 19.

<sup>37</sup> For those not inclined to accept Cartesian dualism, this argument thus also lends support to the view that acts are not individuated by their causes. For a good argument that events are not individuated by their causes, see John Martin Fischer, "Responsibility and Control."

<sup>38</sup> I am grateful to William Hasker for calling my attention to the need to address the issues in this section.

ministically. Since, as the preceding arguments show, it is possible to build FSCs in which the victim really doesn't have any alternative possibilities, these arguments also show that an agent's acting indeterministically doesn't entail his having alternative possibilities available to him. It is therefore possible to hold that moral responsibility requires indeterminism, as van Inwagen says it does, without also having to hold that moral responsibility requires having alternative possibilities.

These conclusions force a certain expansion in the notion of libertarianism. In the past, libertarianism has generally been taken to come in just one kind. It's been understood as the view "[that] people are free and responsible and, a fortiori, that the past does not determine a unique future,"<sup>39</sup> that is, that people aren't causally determined to do the acts which they do freely and for which they are morally responsible. And this has usually been thought to include a commitment to some version of PAP. But if an agent's acting indeterministically does not entail his having alternative possibilities for action, then someone who accepts this description of libertarianism need not also accept PAP. Even if moral responsibility entails indeterminism, it isn't the case that an agent's acting indeterministically entails his having alternative possibilities available to him. Consequently, those who hold the position that "people are free and responsible and . . . that the past does not determine a unique future" aren't compelled by that position to accept some version of PAP. They might still want to maintain PAP for independent reasons, but they aren't committed to PAP by their libertarianism.

Therefore, there are at least two species of libertarianism. Each accepts that people are responsible and that "the past does not determine a unique future," but one accepts and the other rejects PAP.

For the same reasons, as I said at the beginning, it seems to me that the battle line between compatibilism and incompatibilism, which has focused on alternative possibilities, has been misdrawn. Incompatibilism is itself compatible with the rejection of PAP; rejecting PAP doesn't put one in the compatibilist camp.

In my view, both these last points should come only as good news to libertarians. If I am right, then a successful argument against PAP, which has looked like one of the best weapons compatibilists have, is not sufficient to refute either incompatibilism or libertarianism.

---

<sup>39</sup> Robert Audi, *Dictionary of Philosophy* (Cambridge: Cambridge University Press, 1995), p. 281.

ALTERNATIVE POSSIBILITIES AND ULTIMATE RESPONSIBILITY:  
A SUGGESTION

It seems to me worth asking at this point why so many philosophers share van Inwagen's intuition that moral responsibility requires the ability to do otherwise. I don't have an answer to this question which I'm prepared to argue for, but I do want to make a suggestion.

There is something right about the claim made by the proponents of the flicker of freedom defense of PAP, that there is an important difference between an agent's doing an act on his own and his doing it because he is caused to do so by an external intervener. As I have been at pains to argue, the difference is not a difference between different actions the agent does, as the flicker of freedom proponents suppose. Rather, the difference has to do only with *how* the agent does what he does. Even if the victim in a FSC has the same act of will *W* in the actual and the alternative sequence, there nonetheless remains a difference in *how* the victim wills what he does. He is causally determined to an act of will *W* in the alternative sequence, but not in the actual sequence. In the alternative sequence, the ultimate cause of what the victim wills is the intervener; in the actual sequence, it is the victim himself.

Kane has argued that what matters to philosophers who value indeterminism as a requirement for moral responsibility is just our having ultimate responsibility for what we do.<sup>40</sup> If causal determinism is right, then, in some sense which matters to many philosophers, we are not ultimately responsible for what we do. Kane's point seems to me insightful and important, and I am inclined to think that, *mutatis mutandis*, a similar point can be made about the value placed by many philosophers on our ability to do otherwise as a requirement of moral responsibility. If a person *P* does have ultimate responsibility for what he does, then what *P* does is up to *P*, not someone or something else. And if it is up to *P*, then, it seems reasonable to suppose, *P* might do otherwise. On the other hand, if *P* does not have the ability to do otherwise than he does, then it seems as if the ultimate responsibility for what *P* does isn't vested in *P*.

But what the Frankfurt-style counterexamples show, in my view, is that ultimate responsibility and the ability to do otherwise can come apart. Jones in (*G*) wills as he does only because of what he himself believes and desires, and he would will in this way even if there were no counterfactual intervener in the story. He himself, his intellect and will, are the ultimate cause of what he does, and nothing else at all. Consequently, our intuitions are strongly on the side of holding him responsible for his act. If

---

<sup>40</sup> Kane, *The Significance of Free Will*, p. 4 and passim.



his intellect and will are the ultimate cause of his act, he is also ultimately responsible for it.

In FSCs, alternative possibilities are removed because there is an intervener who might, but in fact doesn't, intervene. Controversial though FSCs are, they are the least contentious way to show the separation between ultimate responsibility and the ability to do otherwise. Various philosophers have also argued that there are cases in which something internal to the agent himself might leave him ultimately responsible but without alternative possibilities. An agent might act indeterministically, there might be no cause for his act outside his own intellect and will, and yet he might not have more than one option for action available to him because all the other options are unthinkable for him.

There are many examples in narratives and in the philosophical literature illustrating this point. My own favorite is the story of Ruth and Naomi. Heart-sick, poor, and defeated, Naomi tries to send her daughters-in-law back to their families before she herself turns to a life of begging. One daughter-in-law kisses Naomi and leaves. The other one, Ruth, says, "Entreat me not to leave you or to return from following after you, for where you go, I will go, and where you live, I will live. Your people shall be my people, and your God my God. Where you die, I will die, and there will I be buried. The Lord do so to me and more also if anything but death part you and me" (Ruth 1: 16–17). Can we really suppose that for the woman who made such a speech leaving Naomi was a real alternative possibility? If the passionate speech makes anything clear, it's that leaving Naomi was unthinkable for Ruth. And yet she also seems responsible for what she does – not only responsible but praiseworthy.

Here's a more customary philosophical example. You offer me a nickel to cut my daughter up into little pieces. I love my daughter very much; nickels have their usual value; and there is nothing special about your offer. My accepting your offer won't save the world from terrorists, for example. I can't imagine accepting your offer. But that's because I can see that it's such a bad offer. I lose what is infinitely valuable to me and gain what I value almost at nothing. As long as I have these beliefs and desires, I *couldn't* accept your offer. And yet I see no reason to suppose I'm not responsible for my act of refusing it.

For the same sort of reason, medieval philosophers such as St. Anselm thought that the redeemed in heaven could act with free will (in fact, with the most perfect free will possible) even though it was not possible for them to will anything evil.<sup>41</sup>

---

<sup>41</sup> Failure to come to grips with the point at issue here has also marred some attempts to show that the traditional concept of God is incoherent. For example, some philosophers have supposed that there is something incoherent in the very notion of a perfectly good

But in all these cases, we can suppose, the agent is the ultimate source of her action. The ultimate cause of her action is found in her intellect and will; she does what she does only because of her own beliefs and desires, and there is no other cause of what she does. She is therefore ultimately responsible for what she does. Nonetheless, it isn't necessary for her to have ultimate responsibility that she have alternative possibilities available to her. Her intellect and will might be such that all options but one are unthinkable for her.

So I think Kane is making an important point when he claims that what matters to philosophers who tie responsibility to indeterminism is that an agent be ultimately responsible for his acts. But unless we focus carefully on FSCs or on cases of the sort I've just been describing, we can suppose that having the ability to do otherwise either is or is coextensive with having ultimate responsibility. If this is correct, it helps to explain both why so many philosophers have a strong intuition that PAP is true and also why its rejection should not in fact concern them. If what is valuable about having the ability to do otherwise is that we are ultimately responsible for our acts when we have that ability, then the rejection of PAP isn't worrisome. What the counterexamples to PAP show is only that ultimate responsibility and the ability to do otherwise can dissociate. What they don't show is that moral responsibility and indeterminism can dissociate. Nothing in any of these counterexamples to PAP undermines the intuition that for an agent to be morally responsible, he himself, his beliefs and desires, have to be the ultimate cause of what he does.<sup>42</sup>

*Department of Philosophy*  
*St. Louis University*  
*St. Louis, MO 63103*  
*USA*

---

agent who has his goodness essentially, since moral goodness requires free will, and an essentially perfect agent has to will only the good in every possible world in which he exists. But this argument supposes without reflection that freedom of the will either is or presupposes the ability to do otherwise.

<sup>42</sup> I'm grateful to John Martin Fischer, William Hasker, Norman Kretzmann, Timothy O'Connor, Lee Overton, William Rowe and members of the St. Louis Autonomy Reading Group – Joel Anderson, Lawrence Davis, Sigurdur Kristenssen, and Thad Metz – for helpful comments on an earlier version of this paper.