

# Freedom, Responsibility, and Frankfurt-style Cases

Bruce Macdonald

University College London

MPhilStud – Masters in Philosophical Studies

## **Declaration**

I, Bruce Macdonald, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Signed:

## Abstract

In this thesis I consider an argument against the claim that an agent is responsible for what they have done only if they could have done otherwise. Frankfurt-style cases are proposed as scenarios in which an agent is responsible for what they have done, despite having been unable to do otherwise. A successful Frankfurt-style case would render the question of the compatibility of the *ability to do otherwise* and determinism or indeterminism irrelevant to the question of the compatibility of *responsibility* and determinism or indeterminism. My aim is to assess whether this style of argument succeeds.

I begin by considering a strategy employed by some 'leeway compatibilists' who have argued, via a modified conditional analysis of the ability to do otherwise, that an agent in a Frankfurt-style case could, in fact, have done otherwise in some relevant sense. I argue that these views fail to establish that the agent could have done otherwise in a sense relevant to accounting for that agent's responsibility. I suggest that, for all that these views show, Frankfurt's challenge may stand against leeway compatibilism.

I go on to argue that, insofar as Frankfurt-style cases are proposed to count against 'leeway incompatibilism', determinism must not be assumed, and the counterfactual intervener or intervening mechanism must be equipped to pre-empt the agent's acts of will. I suggest that no dialectically effective Frankfurt-style case can be constructed which would show that the agent could not have done otherwise, in some relevant sense, if it is granted that the agent has the power to determine, without prior determination, their own acts of will. Leeway incompatibilism must be rejected only if there is independent reason to suppose that this ability is unnecessary for responsibility. I conclude that Frankfurt-style cases, in isolation, do not count decisively against leeway incompatibilism.

## Contents

<b>Chapter 1 – Introduction</b> .....	6
1. Freedom, Responsibility, and ‘Leeway’.....	6
2. Causal Determinism and the Problem of Freedom.....	6
(a) Incompatibilism.....	7
(b) Compatibilism.....	9
3. The Consequence Argument.....	10
4. Is Leeway Necessary? – ‘Source Theory’.....	11
5. PAP & Responsibility.....	13
6. Frankfurt’s Strategy and ‘Could Have Done Otherwise’.....	14
<b>Chapter 2 – Leeway Compatibilism and Masked Abilities</b> .....	19
1. The Simple Conditional Analysis.....	19
2. Dispositions and Simple Conditionals.....	19
3. Finkish Dispositions and Lewis’s Modified Conditional Analysis.....	21
4. Masking.....	22
5. Abilities, Conditionals, and Frankfurt-style Cases.....	26
6. Finkish Abilities.....	26
7. Masked Abilities.....	28
8. General and Specific Abilities.....	29
9. In What Sense Was Jones ‘Able’ to Have Acted Otherwise?.....	31
10. Is This Sort of Ability Relevant to PAP?.....	34
11. Significance of General Abilities for the Actual Sequence.....	36
12. The Ability to Try.....	38
<b>Chapter 3 – Leeway Incompatibilism and the Power to Determine One’s Will</b> .....	42
1. Two Kinds of Intervention.....	42
2. The Power to Predict and Pre-empt Activity of the Agent’s Will.....	43
3. The Dilemma Defence of Leeway Incompatibilism.....	45
4. Responding via the ‘Deterministic Horn’.....	46
5. Responding via the ‘Indeterministic Horn’.....	48
6. Leeway Incompatibilist Theories of Freedom.....	48
7. Prior-sign Intervention and Event-causal Views.....	50
8. Blockage Cases.....	51
(a) Hunt’s Case.....	51

(b) Mele & Robb's Case.....	52
9. Flickers of Freedom.....	53
10. Stump's Case.....	56
11. Agent-Causation and the Power to Determine One's Own Will.....	58
12. Conclusion.....	61
Bibliography.....	63

## **Chapter 1 – Introduction**

### 1. Freedom, Responsibility, and ‘Leeway’

When I deliberate about what to do now, it seems to me that I could do various things; I could, for example, continue writing, I could go to the cinema, or I could go swimming. It seems to me that I am *free* to do any one of these things.

A few moments ago, I decided to continue writing, and did so. It seems to me that I am responsible for having done so; some kind of response towards me – of, say, approval, disapproval, or indifference – is warranted in virtue of my having done so. A natural thought is that any such response is warranted only insofar as I could have done otherwise. If deciding to continue writing, and doing so, were the only things that I could have done, then it is not obvious that approval or disapproval, say, are warranted in virtue of my having done so. This might be explained by the thought that, if I could not have done otherwise, then I did not play the kind of role in determining my own behaviour which would make it *mine* – the kind of thing for which *I* am responsible.

This is, I think, a natural way in which to understand what it is to be a free agent, and the relation between an agent’s freedom and their responsibility for what they have done. Firstly, we are free only if we could have done otherwise; call this the ‘*leeway* condition’ for freedom. And, secondly, we are (directly) responsible for what we have done only if we did it freely – that is, only if we could have done otherwise; call this the ‘*leeway* condition’ for direct responsibility.<sup>1</sup>

### 2. Causal Determinism and the Problem of Freedom

I will call *causal determinism* the thesis that a full description of a past state of the world, in conjunction with the laws of nature, entails every fact about present and future states of the world.<sup>2</sup> If this thesis were true, then given the actual past and the laws of nature, there would be only one possible future; those events which are entailed by the actual past and the laws of nature could not fail to occur.

---

<sup>1</sup> Direct responsibility may be distinguished from derivative responsibility; it is less plausible that ‘leeway’ is necessary for derivative responsibility (see §5).

<sup>2</sup> See Fischer, 1994, p. 9

### (a) Incompatibilism

An *incompatibilist* claims that at least one condition, the satisfaction of which is necessary for an agent's freedom (and perhaps, therefore, for their responsibility for what they have done), cannot be satisfied given the truth of the thesis of causal determinism.<sup>3</sup>

One motivation for incompatibilism is the thought that if causal determinism obtains, no agent could ever satisfy the leeway condition for freedom.<sup>4</sup> If it is true that, given a full description of the past, in conjunction with the laws of nature, there is only one possible future, then when I decided to continue writing, and did so, it would seem that I could not have done otherwise; the events of my deciding and acting as I did were entailed by facts about the past and the laws of nature over which I had no control.

Here, then, is a simple argument for what I will call *leeway incompatibilism*: the view that freedom (and, perhaps, responsibility) is (are) incompatible with causal determinism, because no agent could satisfy the leeway condition under causal determinism.

1. We do what we do freely only if we could have done otherwise (and, we are responsible for what we have done only if we did it freely).<sup>5</sup>
2. We could have done otherwise only if, given a full set of facts about the actual past, and the actual laws of nature, up until the moment of one's decision, more than one set of future events is causally possible.
3. If the thesis of causal determinism were true, then given the actual past and the actual laws of nature only one set of future events is causally possible; so, no agent could have done otherwise, and therefore no agent is free in doing (or responsible for having done) what they do (or have done).

According to incompatibilism, it is necessary, at least for freedom, and perhaps also for responsibility, that one's doings (or perhaps their immediate causal antecedents) are undetermined by anything beyond one's control.<sup>6</sup> Therefore, the thesis of causal determinism must be false; indeterminism, of a certain sort, must obtain.<sup>7</sup>

---

<sup>3</sup> As I discuss in §4, some who consider freedom and determinism to be incompatible might consider freedom (at least in this sense) to be unnecessary for responsibility and, perhaps, responsibility to be compatible with determinism (e.g. semi-compatibilists). (Some incompatibilists might consider the truth of other theses of 'determinism' to threaten freedom/responsibility; here I focus on the purported threat posed by the thesis defined in the first paragraph of § 2).

<sup>4</sup> As I discuss in §4, incompatibilism may be motivated by concerns other than the purported threat posed by determinism against agents' satisfaction of the leeway condition.

<sup>5</sup> The bracketed condition is not compulsory for leeway incompatibilists (see fn. 2).

<sup>6</sup> Van Inwagen, 1983, p. 127

<sup>7</sup> Indeterminism, of *some* sort, would obtain if the occurrence of only a single event, somewhere distant from the agent, is undetermined. If an agent's freedom is incompatible with determinism, then it seems clear that their freedom is also incompatible with indeterminism of this sort; nothing, of significance for agential control, is added.

It is not obvious that agents could be free under such indeterministic conditions. It seems to be required, for freedom, that an agent has control, at or immediately prior to some time ( $t'$ ), over whether or not they do something (say, make a decision to perform act A) at this time. If the event of one's deciding, at time  $t$ , to do A, was undetermined, there are possible worlds in which conditions are precisely as they were prior to time  $t$ , in which one does not decide, at  $t$ , to do A. The role played by the agent, at or immediately prior to  $t$ , is not (or, at least not obviously) different in these different worlds. There is, then, reason to suppose that, if the indeterministic conditions, which incompatibilists deem necessary for freedom, obtain, then it is merely a matter of chance whether or not the agent decides, at  $t$ , to do A. And, if so, it would seem to be beyond the control of the agent whether they decide, at  $t$ , to do A, whether they decide otherwise, or whether they make no decision.<sup>8</sup>

There is, then, reason to suppose that freedom is incompatible with the truth of the thesis of causal determinism, *and* that freedom is incompatible with the falsity of this thesis. On the assumption that the thesis is either true or false, these options are exhaustive; so, there is reason to suppose that freedom is impossible, in some sense. But, freedom nonetheless seems to exist; as I have stated, it seems to me that I am now free to do any one of various things. The apparent inconsistency of these claims might be called the 'problem of freedom'.<sup>9</sup>

*Libertarian incompatibilists* propose the following solution. They argue in favour of the following claims: (i) freedom is incompatible with the truth of the thesis of causal determinism, (ii) freedom is compatible with the falsity of this thesis of determinism, and (iii) the relevant sort of indeterminism obtains, such that some agents have freedom. Libertarians face the challenge of developing a plausible account of freedom for which claims (i) and (ii) are true. I discuss some of these views in the third chapter.<sup>10</sup> Libertarians also face the challenge of defending (iii), which seems to be an empirical claim.

Incompatibilists must accept some version of (i).<sup>11</sup> I will call *hard determinists* those incompatibilists who also accept some version of (ii), but who deny (iii) because they claim that the relevant thesis of determinism is true. I will call *hard incompatibilists* those who deny any version of (ii). Both hard determinists and hard incompatibilists deny that any agent actually has freedom; hard incompatibilists deny that any agent could, in some sense, have freedom. Both must claim that our apparent experience of freedom is merely apparent. Both may claim that no agent is responsible for what they do. Alternatively,

---

Undetermined events *at the locus of the agent's actions* seem to be required (but, might not be sufficient) (see, for example, van Inwagen, 1983, p. 126).

<sup>8</sup> Van Inwagen has called this the 'Mind argument' – see his 1983, pp. 126 – 152.

<sup>9</sup> See Van Inwagen, 2002, p. 169

<sup>10</sup> I do so with reference to 'Frankfurt-style cases'; see §4 and onwards.

<sup>11</sup> I add 'some version' because the relevant thesis of determinism may be disputed.

both may argue (as I discuss in §4) that freedom is not required for responsibility, and that responsibility is compatible with determinism, or indeterminism, or both.

### (b) Compatibilism

*Compatibilists* propose to solve the problem by accepting that all conditions which are necessary for freedom may be satisfied, given the truth of the thesis of causal determinism; that is, they deny (i) and claim that, even if the thesis of causal determinism is true, some agents may have freedom.<sup>12</sup>

Compatibilists face the challenge of developing a plausible account of freedom, according to which an agent may be free, compatible with the truth of the thesis of causal determinism.

Some compatibilists endorse what I will call *leeway compatibilism*; they accept the first premise of the above argument for leeway incompatibilism – that satisfaction of the leeway condition is necessary for freedom (and, perhaps, for responsibility) – but deny the second premise of this argument.<sup>13</sup> They argue that, when assessing what an agent could have done, it is legitimate to look to possible worlds in which facts about the past, or the laws of nature, differ from those which obtain in the actual world.

A popular leeway compatibilist strategy has been to analyse an agent's ability to have done otherwise in terms of simple counterfactual conditionals; one 'could have done otherwise' is analysed as 'one would have acted otherwise, if one had decided (willed/chosen/tried, etc.) to act otherwise'.<sup>14</sup> If this analysis is correct, then we may say that I could have refrained from raising my arm, even if I did not actually do so, and causal determinism obtains. For example, I am such that, if I decide to raise my arm, and nothing constrains my doing so, I would raise my arm. And, similarly, if I decide to refrain from raising my arm, I would refrain. It is true of me that, if I were to have decided to refrain from raising my arm, I would have refrained. It is consistent with the thesis of causal determinism that, had a different set of past events occurred, a different set of future events would have occurred. And, if I had decided to refrain (given that I did not actually decide to refrain), at least one fact about the past would have been different.

Analyses of this sort have faced various difficulties. In the second chapter, I discuss recent attempts to resurrect this strategy. I do so with particular reference to the implications of these views for 'Frankfurt-style cases', which I introduce in §4 of this chapter. These are cases which purportedly undermine leeway compatibilist and leeway incompatibilist accounts of freedom as accounts of the kind of control necessary for responsibility.

---

<sup>12</sup> Compatibilists might, but need not, deny that freedom is compatible with indeterminism of any sort.

<sup>13</sup> As I go on to discuss in §4, other 'non-leeway' versions of compatibilism are available.

<sup>14</sup> Ayer, 1954, pp. 15 – 23 (I discuss some epistemological and metaphysical motivations for such analyses in C.2, §2).

### 3. The Consequence Argument

If causal determinism obtains, then my earlier decision and other actions were entailed by facts about the past and the laws of nature. For me to have decided or acted otherwise than I did, different prior events must have occurred, or different laws of nature must have obtained. One who wishes to maintain that I *could*, at or just before the time of my decision or action, have decided or acted otherwise, given causal determinism (as a leeway compatibilist must), is committed to the claim that I could, then, have behaved such that the past, or the laws of nature, would have been different. But, this seems to be absurd. Firstly, it seems that no person can, at some time ( $t$ ), act such that some fact about the past, prior to  $t$ , would be rendered false.<sup>15</sup> Secondly, it seems that no person can alter the laws of nature such that a proposition entailed by a law of nature would be rendered false.<sup>16</sup> And, it is plausible that (a) one's powerlessness over the past and the laws, and (b) one's powerlessness over the fact that the conjunction of the past and the laws entailed how one decided and acted, together entail (c) one's powerlessness over how one decided and acted.<sup>17</sup> We have reason, then, to suppose that, if causal determinism obtains, I *could not* have decided or acted otherwise than I did.

This is an informal expression of the 'consequence argument', which some take to count decisively against leeway compatibilism.<sup>18</sup> There is much debate over the soundness of this argument. David Lewis, for example, has argued that it is unsound.<sup>19</sup> He distinguishes between (i) being able to perform some act that is itself, or is a cause of, a law-breaking event (i.e. being able to render false a conjunction of the laws of nature ( $L$ ), in a strong, or causal, sense), and (ii) being able to perform some act such that, if one were to perform it, a law would not have obtained, but not because one performed this act (i.e. being able to render  $L$  false in a weak, or non-causal, sense). According to Lewis, the consequence argument leaves open the possibility that we are able to render  $L$  false in this weak sense. It is true, for example, that if, contrary to fact, I were to have acted otherwise at  $t$ , then (assuming the fixity of the past) a law-breaking event – a 'divergence miracle' – which is incompatible with the actual laws, must have occurred at some point prior to  $t$ .<sup>20</sup> Hence, he claims, leeway compatibilism may stand.

Note that, if some conditional analysis of 'could have done otherwise' were correct, then the inference from (a)-(c) would be invalid. Suppose that I performed some act ' $A$ ' at  $t$ . If, contrary to fact, I were to have decided before  $t$  to perform some other act ' $B$ ' at  $t$ , I would have performed this act at  $t$ ; so, on a conditional analysis I *could* have  $B$ -ed at  $t$  and thereby done otherwise than  $A$  at  $t$ . Therefore, I would not

---

<sup>15</sup> Call this the 'principle of the fixity of the past'.

<sup>16</sup> Call this the 'principle of the fixity of the laws'.

<sup>17</sup> This rule of inference has been termed the 'principle of the transfer of powerlessness'; it alleges that the property of 'being powerless over whether or not  $p$ ' is closed under logical entailment.

<sup>18</sup> See van Inwagen, 1983, p. 16, and Ginet, 1980, pp. 171 – 186

<sup>19</sup> See Lewis, 1981 (reprinted in Watson, ed. 2003), p. 123 (a parallel argument might be given with regard to the fourth premise, but Lewis does not to make this argument – see his 1979, pp. 456 – 457)

<sup>20</sup> Fischer, 1994, p. 70

have been powerless over what I did at *t*. But, I would have been powerless over facts about the past and the laws of nature, and over their entailing what I did at *t*. As I have mentioned, however, analyses of this sort face difficulties, which I will go on to discuss.

#### 4. Is Leeway Necessary? – ‘Source Theory’

The consequence argument does not directly refute the leeway compatibilist’s view, but it does place a burden upon leeway compatibilists; they must develop an independently convincing account of an agent’s ability to do otherwise, given causal determinism. There is, however, reason to think that debate over the consequence argument does not get to the heart of the matter, at least with regard to the conditions necessary for responsibility. The question of whether leeway is necessary for responsibility, and perhaps for freedom, threatens to undercut this debate.

Various arguments challenge some aspect of ‘leeway theory’ (i.e. the view that leeway – the ability to do, or have done, otherwise – is necessary for freedom and, perhaps, responsibility).<sup>21</sup> In what follows, I wish to consider in detail one such challenge; this is due to Harry Frankfurt, who has argued that satisfaction of the leeway condition is not necessary in order for us to be responsible for what we have done.<sup>22</sup> If his argument is sound, then we should reject leeway incompatibilist and leeway compatibilist accounts of freedom as accounts of the kind of control necessary for responsibility. This would permit agnosticism about the soundness of the consequence argument – even if we could not, under causal determinism, have done otherwise than behave as we did, this need not count against our responsibility for what we have done, because we need not have been able to do otherwise in order to be responsible for what we have done.

It is natural to suppose that something that I apparently did (e.g. my performance of a bodily movement) was genuinely *mine* only insofar as *I* was its source or origin, in some sense; that I brought it about, via some power of mine. This thought motivates what we might call the ‘*source* condition’ for responsibility: one is responsible for what one has done only if one was its source, in some sense. One might satisfy the source condition if one satisfies the leeway condition; for example, if I decided to do something, selecting it from among open alternatives, and thereby determining that I did this thing rather than something else, I might be the source of what I did. But, it is not obvious that satisfaction of the leeway condition is *necessary* for satisfaction of the source condition.<sup>23</sup> For example, one might be free to do as one wills, insofar as no external constraint prevents one from so acting, even if one lacked leeway; one’s will could still be the source of one’s behaviour, and determine what one does, even if one could not have willed or

---

<sup>21</sup> E.g. Strawson (P.F.) 1962, and Dennett, 1984, pp. 131 – 152. A full defence of leeway theory would have to engage with these arguments.

<sup>22</sup> Frankfurt, 1969 (reprinted in his 1988), pp. 1 – 10

<sup>23</sup> Kane has argued that leeway is necessary for source-hood; see his 1996, pp. 60 – 78. I take no stand on this issue.

acted otherwise – in such cases, it would be true that one *did* will in this way, and one did act in accordance with, and on the basis of, one's will.

If one accepts that one might be the source of what one has done (in a sense which could ground one's responsibility for what one has done) even if one could not have willed or acted otherwise, then one accepts what I will call a 'source theory' of the kind of control required for responsibility. Frankfurt argues for the rejection of leeway theory and the acceptance of source theory. But, source theory is consistent with either the incompatibility or compatibility of causal determinism with the kind of control required for responsibility. His argument is not a direct argument for the compatibility of determinism and responsibility. Whether or not one could be the source of what one has done, if causal determinism obtains, depends upon the kind of 'source' which one must be in order to be responsible for what one has done.

For example, I would be the source of what I have done, in one sense, if my act of will were the immediate and determining antecedent of my action. On this view, satisfaction of the source condition would be compatible with causal determinism; my act of will may play this role even if it is itself the determined effect of some prior event which has its causal origin outside of me. But, there is some intuitive support for the thought that, if this were the case – if all of my actions had this kind of causal history – I would never be the *genuine* source of what I do; for genuine source-hood, *I* – or at least some event within me – must be the ultimate source or origin of my action: there must be no conditions, sufficient for my action, which are independent of me, the agent. And, if this kind of 'ultimate' source-hood is necessary for satisfaction of the source condition, then responsibility would be incompatible with causal determinism.

Frankfurt's argument is, then, consistent with '*source incompatibilism*': the view that no agent satisfies the source condition if causal determinism is true. This view is resisted by those who argue that causal determinism does not threaten satisfaction of the source condition, in the sense required for responsibility; those who take this view endorse '*source compatibilism*'. Source compatibilists face the challenge of showing a relevant distinction between acting under causal determinism – circumstances in which they claim that agents may be responsible for what they do – and acting under certain kinds of manipulation – circumstances in which it seems that agents are not responsible for what they do.<sup>24</sup> I take no stand on the relative merits of source incompatibilism and source compatibilism. My concern is with the success of Frankfurt's argument against leeway theory.

Accepting a source theory of the kind of control required for responsibility does not entail accepting a source theory of freedom; one might retain a leeway theory of freedom, and merely reject the claim that

---

<sup>24</sup> See Kane, 1996, pp. 64 – 71 and Pereboom, 2007, pp. 93 – 98

we are responsible for what we have done only if we did it freely (i.e. could have done otherwise). Some who accept a source theory of the kind of control necessary for responsibility might claim that we are 'free', at least in any sense worth wanting, when we possess this kind of control. It is not obvious, however, that even if the kind of control required for responsibility is distinguishable from the ability to do otherwise, the former would be the *only* kind of 'freedom' worth wanting. Fischer, for example, has argued for a leeway incompatibilist theory of freedom, combined with a source compatibilist account of the kind of control required for responsibility; he calls his view 'semi-compatibilism'.<sup>25</sup>

## 5. PAP & Responsibility

Frankfurt's argument is directed against what he called the 'principle of alternate possibilities', abbreviated as 'PAP'. According to PAP, 'a person is morally responsible for what they have done only if they could have done otherwise'.<sup>26</sup> This is similar to what I have called the 'leeway condition for responsibility'. Note that Frankfurt's PAP proposes a necessary condition for our *moral* responsibility for things which we *do*. Insofar as we perform actions and bring about states of affairs, we engage in 'positive agency' – we do things, for which we may be responsible. But, instances of positive agency do not exhaust those things for which we may be responsible; this includes instances of 'negative agency' – things which we do not do. In what follows, I will focus on instances of responsibility for positive agency; I acknowledge that nothing follows, without further argument, with regard to the necessary conditions for responsibility for negative agency.

As I will use the term, we are *responsible* for performing actions or bringing about outcomes in virtue of which we, the agent, warrant a response of some sort.<sup>27</sup> We may be *morally* responsible for those performances of which *moral* responses, of some sort, are warranted; these include those things for which we are morally blameworthy, and those for which we are morally praiseworthy. But, perhaps one might be blameworthy or praiseworthy, for what one has done, in a non-moral sense which is not merely superficial. For example, we might be worthy of response in virtue of certain kinds of achievements which are recognisably ours; we may be worthy of praise in virtue of, say, sporting achievement, but such praise need have no moral content. And, some kinds of doings or non-doings of ours are unremarkable, and neither blameworthy nor praiseworthy; a response towards us is warranted, in virtue of their being ours, but this response is morally-neutral.

One might wish to include such doings or non-doings within the category of things for which we are morally responsible; perhaps they are potentially morally significant events. Alternatively, one might wish

---

<sup>25</sup> Fischer, 1994, p. 178

<sup>26</sup> Frankfurt, 1969, p. 1

<sup>27</sup> This would exclude, then, cases of superficial, or merely causal, responsibility; in these cases, the event is the object of the response, and not the agent – hence, non-agents can be responsible for events in this sense (See Wolf, 1990, p. 40).

to retain a narrower category of events for which we are morally responsible, and allow that we are responsible for some things for which we are not morally responsible. In what follows, my use of ‘responsibility’ is intended to include such events for which we might not be morally responsible, in this narrow sense, although I take no stand on the correct usage. I focus on broader responsibility because narrow focus on the presuppositions of specifically *moral* praise and blame might distract from what I take to be of primary significance; that is, which kind of control over what we do is necessary in order for what we do to be *ours*, on which our responsibility, and moral responsibility, for what we do seems to depend.

PAP specifies a necessary condition for *direct* responsibility, and not for derivative responsibility. For example, the principle would not be threatened by a case of the following sort: ‘at the time of their accident, it was genuinely beyond the drunk driver’s power to have done otherwise than crash their car, but it does seem, nonetheless, that they are responsible for having done so’. The drunk driver may be derivatively responsible for having crashed their car, in virtue of their direct responsibility for having become drunk and driven their car.<sup>28</sup> This may be so, consistent with PAP, insofar as at some earlier stage this individual could have done otherwise than make it the case that, at some later time they would be unable to do otherwise than behave as they did. In this case, the agent need not have become drunk. Note, however, that PAP specifies a necessary condition, not a sufficient condition, for responsibility; so, even if this person had the kind of control over their behaviour which is necessary for direct responsibility, they might have failed to satisfy a further necessary condition; for example, it might not have been reasonable to have expected them to be aware of the risks inherent in behaving as they did.<sup>29</sup>

## 6. Frankfurt’s Strategy and ‘Could Have Done Otherwise’

Frankfurt argued against PAP via a proposed counterexample. He suggested that there is at least one possible scenario in which an agent has done something, it is false that the agent ‘could have done otherwise’, and, because the factor which renders the agent ‘unable to have done otherwise’ is distinguishable from the factor which explains why the agent behaved as they did, the agent is nonetheless responsible for what they have done. In Frankfurt’s examples, the agent behaved as they did because they willed, in the normal way, to do so, and were unconstrained in doing as they willed; they, purportedly, could not have done otherwise, because of the presence of some overdetermining intervener or mechanism, which ensures that they would have behaved as they did, whether or not they willed, in the normal way, to do so.

---

<sup>28</sup> See Mele, 2006, pp. 84 – 85

<sup>29</sup> They might, say, have drunk alcohol freely but unwittingly; certain sorts of responses to them, in virtue of their behaviour might, therefore, be unwarranted.

If an agent performs an action and brings about an outcome, and could have successfully brought about an alternative type-outcome, then they would satisfy PAP with regard to what they have done.<sup>30</sup> But, it seems to be false that an agent could be responsible for what they have done only if they could have successfully brought about an alternative type-outcome. There may be cases in which the type-outcome was overdetermined; the type-outcome which one actually brought about via one's action would have occurred anyway, whether or not one had acted as one did.

For example, suppose that an agent, Jones, decides to perform, and successfully performs, some action, thereby bringing about some outcome. He is confronted by, say, a bowl of blue sweets and a bowl of red sweets and is encouraged to eat whichever he likes; after deliberating about which, if any, to select, he decides to eat a blue sweet, and successfully does so. Suppose there are no further morally significant features of the case – eating sweets of either sort is a morally neutral action, but it is nonetheless *Jones's* action. And, Jones could have done otherwise than behave as he did, in whichever sense one takes to be relevant; one may suppose that he could have done otherwise on this precise occasion, given the actual past and the laws of nature, or merely that he could have done otherwise, had the past or the laws been different. So, it is uncontroversial to consider Jones to be responsible for what he has done.

Now, suppose a relevantly similar scenario, but one in which, prior to Jones's decision, an 'unsophisticated intervener' – Dr. Green – was present. Dr. Green was keen that Jones consume at least one blue sweet, and only blue sweets. But, he was also lazy, and did not wish to expend any unnecessary effort. His presence rendered true the following counterfactual.

(a) If Jones were to have decided to eat a red sweet, or no sweets, and made an overt attempt to eat a red sweet, or to refrain from consuming any sweets, Dr. Green would have prevented him from successfully doing either, and would have force-fed him a blue sweet.

Jones actually decided to eat a blue sweet, and ate one; Dr. Green had no cause to intervene, so he did nothing – he played no role in bringing about this outcome. So, it seems uncontroversial to judge Jones to be responsible for having eaten a blue sweet, despite it being true that he could not have avoided consuming, somehow, a blue sweet. This seems, then, to be a counterexample to the claim under consideration: that an agent is responsible for what they have done only if they could have successfully brought about an alternative type-outcome.

This example leaves PAP intact, unless we suppose that one could have done otherwise, in the sense relevant to one's responsibility for what one has done, only if one could have successfully brought about

---

<sup>30</sup> If the actual outcome involved, say, 'Jones consuming a blue sweet', then one could have brought about an alternative *type*-outcome by making something other than 'Jones consuming a blue sweet' occur. There may be variations among specific instances – *tokens* – of the same type of outcome. There is some controversy over the significance of alternative token-outcomes, actions, or decisions for PAP, or suitable replacement principles; I come back to this in the third chapter. See van Inwagen, 1978, pp. 205 – 221, and Fischer, 1982, pp. 29 – 32.

an alternative type-outcome. And, this supposition is not particularly compelling. In this case, Jones retained the power to have decided otherwise, and to have made an overt attempt to act otherwise; by exercising these powers, he would have made it the case that Dr. Green, rather than Jones himself, brought about this type-outcome. Given that he did not exercise this power, it seems like he brought about this outcome via a free decision and action; his freedom in deciding and acting as he did does seem to be relevant to his responsibility for what he has done. So a plausible version of PAP is not threatened by this case.

'Frankfurt-style cases' have this general structure, but may involve a more sophisticated intervener or mechanism. For example, if an agent performs a bodily action and could have successfully performed an alternative type-action, then they would satisfy PAP with regard to what they have done. But, it is not obvious that an agent could be responsible for what they have done only if they could have successfully performed an alternative type-action. There may be cases in which one's performance of a bodily action was overdetermined; one might have performed a bodily action of the same type (e.g. 'eating a blue sweet') anyway, even if one had not willed on one's own to do so. For example, suppose that, in the scenario described above, we replace Dr. Green with a more sophisticated intervener – Dr. Yellow – who is capable of monitoring Jones's decisions. Dr. Yellow's presence rendered true the following counterfactual.

(b) If Jones were to have decided to eat a red sweet or no sweets, Dr. Yellow would have immediately prevented him from beginning to perform any of the bodily movements (or non-movements) necessary for Jones to eat a red sweet, or to refrain from eating any sweets, and he would, then, have determined that Jones would decide to eat a blue sweet, and succeed in doing so.

Jones actually decided to eat a blue sweet, so Dr. Yellow had no cause to intervene; he did nothing, and played no role in bringing about Jones's subsequent action, and the outcome. So, it seems uncontroversial to judge Jones to be responsible for what he has done.

There are, however, at least two reasons to suppose that this case leaves PAP intact. Firstly, we need not suppose that one could have done otherwise, in a sense relevant to one's responsibility for what one has done, only if one could have successfully performed an alternative type-(bodily)action. In this case, Jones retained the power to have decided otherwise; by exercising this power, he would have made it the case that Dr. Yellow, rather than Jones himself, brought about his performance of this type-action. Given that he did not exercise this power, it seems like he brought about this type-action via a free decision; his freedom in deciding as he did does seem to be relevant to his responsibility for what he has done. And, secondly, it is not clear that, in this case, Jones really could not have performed an alternative type-action, at least in some sense of 'could' which might be relevant to his responsibility for what he has done.

I will discuss the second of these points at length in the second chapter. As regards the first, a yet more sophisticated intervener can be introduced; Frankfurt's most detailed version of the example runs as follows. In the above scenario, Dr. Yellow is replaced by a very sophisticated intervener – Dr. Black – who is capable, somehow, of reliably predicting Jones's decisions. Dr. Black will intervene and manipulate Jones when required (i.e. when Jones would otherwise have decided otherwise), but will do nothing when he predicts that Jones will decide and act as he (Black) wishes (like Green and Yellow before him, Black wishes Jones to consume at least one blue sweet, and only blue sweets). Dr. Black's presence rendered true the following conditionals.

(c) If Jones were about to decide to eat at least one blue sweet, Black would have predicted this and done nothing, thereby allowing Jones to decide on his own to do so, and to act as he decides.

(d) If Jones were about to decide to eat a red sweet, or to eat no sweets, Black would have predicted this and ensured that Jones does not go on to make this decision; he would have pre-empted Jones's decision by determining that Jones decides to, and does, eat at least one blue sweet.<sup>31</sup>

In the actual scenario, Jones decided ('for reasons of his own' – i.e. without Black's intervention) to eat a blue sweet. Dr. Black would have predicted this, and therefore had no cause to intervene; he would have done nothing, and played no role in bringing about Jones's subsequent decision, action, and the outcome. So, it seems uncontroversial to judge Jones to be responsible for what he has done. That is, our judgments about whether or not Jones is responsible for having eaten a blue sweet in this case should be the same as our judgments about whether or not he would be responsible for having done so, had Dr. Black been entirely absent from the scene.<sup>32</sup> And, it has already been granted that Jones would have been responsible for having done so, had Black been entirely absent. So, intuitively, Jones is responsible for having eaten a blue sweet in this case.

In this case, Jones seems to lack the power to have done something which would have made it the case that Dr. Black, rather than Jones himself, brought about his decision, action, and the outcome. Dr. Black's intervention would occur on the basis of his prediction, before Jones could decide, or even begin to decide, what to do. According to Frankfurt, there is no sense in which Jones could have done otherwise in this case, which might be relevant to his responsibility for what he has done. Frankfurt concludes that PAP is false; being able to have done otherwise is not a necessary condition for one to be responsible for what one has done.

There may, however, be some sense in which Jones 'could' have done otherwise, even in this more sophisticated case, which might be relevant to his responsibility for what he has done; in the second chapter, and in the early parts of the third chapter, I consider arguments which purport to defend leeway

---

<sup>31</sup> Perhaps Jones might fail to decide either way, before some specified time. We may suppose that Dr. Black would also intervene in this case, and determine that Jones decides to eat a blue sweet.

<sup>32</sup> Frankfurt, 1969 (in his 1988), p. 7

compatibilism in this way. I then go on to consider a different style of defence of leeway incompatibilism; it may be the case that the structure of the counterexample depends upon assumptions which are unacceptable from the point of view of at least some leeway incompatibilist accounts of freedom.

## **Chapter 2 – Leeway Compatibilism and Masked Abilities**

### **1. The Simple Conditional Analysis**

In the second ‘Frankfurt-style case’ that I presented – the case of Jones and Dr. Yellow – it was proposed that Jones could not have performed an alternative type-action. Given Dr. Yellow’s presence, the following counterfactual is false: ‘if Jones were to have decided (willed/chosen/tried etc.) to perform an action other than of the type ‘eating a blue sweet’, he would have succeeded in performing this action, or would at least have made an overt attempt to perform this action’.

This is an apparently significant result. As I mentioned, some leeway compatibilists have attempted to analyse an agent’s ability to have done otherwise, in terms of simple conditionals: one ‘could have done otherwise’ is analysed as ‘one would have acted otherwise, if one had decided (willed/chosen/tried, etc.) to act otherwise’.<sup>33</sup> Given Dr. Yellow’s presence, Jones could not have done otherwise in this sense; in all cases in which Dr. Yellow is present and Jones decides (wills/chooses/tries etc.) to perform an action other than of the type ‘eating a blue sweet’, he will not successfully complete this action, nor even initiate it (aside from deciding to perform it). Yet, given Dr. Yellow’s actual non-intervention, Jones does seem, nonetheless, to be responsible for what he has done. So, insofar as one accepts that the truth of this simple counterfactual conditional is necessary for the agent in question to be able to do otherwise in the sense relevant to PAP, then one should accept Frankfurt’s conclusion: being able to have done otherwise is not a necessary condition for one to be responsible for what one has done.

Leeway compatibilists need not, however, claim that the truth of this simple conditional is necessary for the agent to be able to do otherwise in the sense relevant to PAP. Recent advances in the analysis of dispositions have prompted the development of more sophisticated conditional analyses of abilities. Applied to the ability to do otherwise, these analyses promise to avoid problems facing the simple conditional analysis, and to show that Frankfurt’s conclusion is unwarranted. In what follows, I outline the motivation for this line of argument by discussing recent debate about the analysis of dispositions; I then consider analogous analyses of abilities, and consider their implications for Frankfurt’s argument.

### **2. Dispositions and Simple Conditionals**

This Frankfurt-style case is structurally similar to thought experiments employed in other contexts, each of which proposes a counterexample against an attempt to analyse some phenomena in terms of

---

<sup>33</sup> See Ayer, 1954, pp. 15 – 23, and Moore, 1911, p. 103

counterfactual conditionals.<sup>34</sup> These examples highlight instances of a general problem with attempts to analyse the truth or falsity of some statement ‘P’ as dependent, at least sometimes, on the truth or falsity of a conditional, of the form ‘if A were to occur, then B would occur’. It is a form of what Shope has called the ‘conditional fallacy’: in some relevant situations, statement P is true, but A’s occurring would cause (or at least be a partial cause of) B’s failure to occur.<sup>35</sup> In the Frankfurt-style case, the truth of A (‘that Jones decides to act otherwise’) triggers Dr. Yellow’s intervention, which ensures the falsity of B (‘that Jones acts otherwise’). The analysans is therefore false, because the consequent may be false while the antecedent is true. And, because the analysandum seems to be true in cases in which the analysans is false, we have reason to agree that the truth of this conditional statement is not necessary for the truth of the analysandum.

Some examples of structurally similar cases can be found in C.B. Martin’s proposed counterexamples against the simple conditional analysis of dispositions. Objects have dispositions; for any disposition, some (perhaps merely hypothetical) conditions stimulate its manifestation. An object’s disposition of fragility, say, is manifested when an object breaks, given certain stimulus conditions – when, for example, it is struck; a fragile object is one which is disposed to break when these conditions obtain. Given a pattern of manifestations under certain stimuli, there seems to be some relation between true disposition ascriptions and true conditionals. Some have proposed to analyse disposition ascriptions in terms of such conditionals, such that to claim that something is fragile, for example, means nothing more than that ‘it would manifest its fragility (e.g. break) given stimulus conditions C (e.g. if it were struck)’.<sup>36</sup>

Martin suggested the following counterexamples to this analysis. Consider the analysis ‘a wire is live (i.e. disposed to conduct electricity) if and only if it would conduct electricity when touched by a conductor’.<sup>37</sup> Suppose that a dead wire is connected to an ‘electro-fink’ – this is a device which ensures that, whenever the wire is touched by a conductor, it would conduct electricity. When the wire is not touched by a conductor it lacks the disposition to conduct electricity, but the counterfactual conditional is nonetheless true of it; if it were touched by a conductor, it would conduct electricity. So, the truth of this conditional, relative to some object at some time, seems to be *insufficient* for the ascription of a disposition to this object. And, suppose that a live wire is connected to a ‘reverse-cycle electro-fink’ – this device ensures that, whenever the wire is touched by a conductor, it would not conduct electricity. When the wire is not

---

<sup>34</sup> See, for example, Nozick, 1981, p. 179 (the ‘grandmother’ counterexample to a simple truth-tracking analysis of knowledge), Lewis, 1987, pp. 277 – 278 (the ‘light-meter’ counterexample to a simple causal theory of perception), and Martin, 1994, pp. 2 – 3 (the ‘electro-fink’ counterexamples to the simple conditional analysis of dispositions).

<sup>35</sup> Shope, 1978, p. 400

<sup>36</sup> See, for example, Ryle, 1959, pp. 117 – 119. This may be motivated by deeper epistemological and metaphysical commitments. An empiricism according to which one has knowledge only of that which can be experienced, and only events (and not properties or substances) can be experienced, might motivate exclusion of dispositions, qua properties, from one’s ontology; stimulus and manifestation events are observable, whereas a disposition, qua property (e.g. ‘causal power’ or ‘potentiality’), is not. In order to make sense of disposition ascriptions, then, one might analyse them as statements which report contingent regularities in observable events, and do not ascribe mysterious categorical properties to their objects (see Mumford, 1998, pp. 26 – 29 & pp. 36 – 44).

<sup>37</sup> See Martin, 1994, pp. 2 – 4

touched by a conductor it has the disposition to conduct electricity, but the conditional statement is nonetheless false of it. So, the truth of the conditional statement seems to be *unnecessary* for the truth of the disposition ascription.

### 3. Finkish Dispositions and Lewis's Modified Conditional Analysis

In Martin's first example, the wire's *lack* of the disposition to conduct electricity is 'finkish'; the fink-device causes the wire to have the disposition to conduct electricity when and only when the stimulus conditions obtain. In the second case, the wire's *possession* of this disposition is 'finkish'; the reverse-cycle fink-device removes the wire's disposition to conduct electricity when and only when its stimulus conditions obtain.<sup>38</sup>

In order to resist these counterexamples, one would have to claim that the addition of a fink-device of either sort makes an *actual* change to the dispositions of the wires, rather than a merely counterfactual change. There are clear examples of objects acquiring or losing dispositions over time; for example, a formerly fragile glass is no longer fragile once it has been melted. But, this change in disposition occurs in virtue of a real change in the internal structure of the glass. And, in Martin's examples, the fink-device would become active – and, perhaps, change the internal structure of the wire – only when the wire is touched by a conductor; when the fink-device is dormant, the internal structure of the wire is unchanged. So, in order to claim that the mere addition (and dormant presence) of a fink-device changes the disposition of the wire, one would be committed to the view that an object's dispositions may change, without it undergoing any intrinsic change.<sup>39</sup>

Against this view, it is plausible to suppose that intrinsically identical objects, subject to the same laws of nature, are disposed alike.<sup>40</sup> David Lewis has argued, on the basis of this supposed intrinsicness of the causal bases of dispositions, in favour of the following, more sophisticated conditional analysis: an object ('O') is disposed to exhibit some manifestation ('M') in conditions ('C') if and only if O has an intrinsic property ('B') such that, if C obtains, and if O were to retain B, then B and C would jointly cause O to M.<sup>41</sup> This provides intuitively satisfying results in both of Martin's examples. If a dead wire is attached to an electro-fink, then the wire would conduct electricity if touched by a conductor, but it lacks any intrinsic property such that, if it were to retain this property when touched by the conductor, this property would be part of the cause of its doing so; on Lewis's analysis, this wire is not disposed to conduct electricity

---

<sup>38</sup> Lewis, 1997, pp. 143 – 144

<sup>39</sup> Roughly, an object's *extrinsic* properties are those that it has in virtue of how it interacts with the world; an object's *intrinsic* properties are those which it has in virtue of the way that it is, in itself.

<sup>40</sup> See Lewis, 1997, p. 148 – This is not obviously true for all dispositions, as there might be extrinsic dispositions; McKittrick (2003) suggests weight, and visibility as possible examples. Even so, this need not count against Martin's argument; there may be constraints on the ways in which an object can extrinsically gain or lose dispositions, which are not met by the mere presence of an electro-fink – see Fara, 2005, pp. 47 – 48.

<sup>41</sup> Lewis, 1997, p. 157

when the electro-fink is dormant. Similarly, a live wire attached to a reverse-cycle electro-fink would not conduct electricity if touched by a conductor, but it does have an intrinsic property which, if it were to retain this property when touched by the conductor, would cause it to do so; on Lewis's analysis, this wire is disposed to conduct electricity when the reverse-cycle electro-fink is dormant.

#### 4. Masking

The fink-devices in Martin's examples cause intrinsic change in the object when and only when the stimulus conditions obtain; in these circumstances, the relevant disposition is lost or gained, and this ensures non-manifestation, given the stimulus conditions, which falsifies the simple conditional. But, other kinds of interference may ensure non-manifestation, given the stimulus conditions, without causing any intrinsic change in the object. For example, suppose that a glass is fragile. This glass is lined with packing material, such that the simple conditional 'if it were struck, it would break' is now false. The process of lining the glass does not cause intrinsic change in the glass, but it does prevent the manifestation of the glass's disposition to break when struck, by blocking the processes which would, in normal conditions, lead from the glass being struck, to it breaking.<sup>42</sup> This case is, then, distinct from those in which the object's having or lacking a disposition is 'finkish'; in this case, the manifestation of the object's disposition is 'masked'.<sup>43</sup>

Masking seems to pose a problem for Lewis's analysis. When it is lined with packing material, the glass retains the intrinsic properties which constituted the causal basis of its disposition to break when struck. But, it would not now break, if it were struck; and, *a fortiori*, this property would not now be part of the cause of its breaking when struck. So, it looks like Lewis's sophisticated counterfactual conditional is false, while the object retains the disposition in question; this seems to be a counterexample – the truth of Lewis's analysis is not necessary for true disposition ascriptions.

Lewis has suggested, in response, that there is no such problem when the stimulus conditions and response are specified correctly.<sup>44</sup> For example, if we define 'poison' roughly, as 'a substance which is disposed to cause death if ingested', then the problem arises; a poison might be ingested, yet fail to cause death because its effects are masked by the ingestion of an antidote. But, according to Lewis, "the specifications of both the response and the stimulus stand in need of various corrections".<sup>45</sup> We should say, for example, 'a substance which is disposed to cause death *in a certain sort of organism* if ingested *without*

---

<sup>42</sup> Johnston, 1992, p. 233

<sup>43</sup> Cases in which an object is, say, non-fragile, but is extrinsically affected such that it *would* break under the relevant conditions are cases of 'mimicking'. These cases challenge the sufficiency of Lewis's analysis for true disposition ascriptions.

<sup>44</sup> Lewis, 1997, p. 153

<sup>45</sup> *Ibid.* p. 153

*its antidote*?. So, the failure of the poison to cause death when ingested alongside its antidote does not show that Lewis's analysis fails for this more clearly specified disposition.

Lewis claimed that these precisely specified dispositions are what we actually express in our ordinary conventional or canonical disposition ascriptions.<sup>46</sup> When we say 'this object is fragile', or 'this object is disposed to break when struck', we are employing a concept with precisely specified conditions of application, the full articulation of which may require detailed conceptual analysis.<sup>47</sup> Such a full articulation of these conditions would exclude *any* cases of masking, and thus ensure that there is, in principle, always some sophisticated conditional which is equivalent to the disposition ascription.

This has some intuitive support. Suppose that Jones claims that 'this glass is fragile'; asked to clarify what he means, he states 'it would break if it were dropped'. But, it is then pointed out to Jones (by a pedantic friend) that this is false, if the glass were lined with protective material, or were dropped from one millimetre above the floor, or were dropped onto a densely carpeted floor.<sup>48</sup> It is plausible that Jones's ascription of fragility was not wrong, but that he was insufficiently precise, or took it to be contextually obvious what he meant; he really meant that, if certain conditions are held fixed – the floor is hard concrete, the glass is on a high table, and it is not wrapped in protective material – the glass would break, if it were dropped.

But, this strategy faces some problems. Jones may rule out *some* potential maskers by being specific about which stimulus conditions he means, but we can imagine other maskers obtaining, consistent with the specified conditions. For example, suppose that the area between the high table and the concrete floor is filled with viscous liquid; the unprotected glass would fall slowly through the liquid and, upon striking the concrete floor, remain intact.<sup>49</sup> So, perhaps he should be even more specific, and add a further condition – that there is no viscous liquid between table and floor – to the precise description of stimulus conditions. But, presumably many disparate potential maskers could be imagined which are not yet ruled out. It is not clear that any finite list could exclude *all* potential maskers.<sup>50</sup> Nor is it clear that there could be any principled basis on which to add such conditions to one's list, other than that those conditions would prevent the relevant manifestation. And, if one were to claim that 'a glass is fragile only if it would break when dropped given conditions in which it would break when dropped', then one's conditional analysis would be trivial and uninformative.<sup>51</sup>

---

<sup>46</sup> Ibid. p. 153

<sup>47</sup> The correct specification of conditions will also be context dependent; a working engineer will have different standards for 'fragility' than a working chemist, say (see Manley & Wasserman, 2008, p. 65, and pp. 71 – 72).

<sup>48</sup> See Manley & Wasserman, 2008, p. 65

<sup>49</sup> Manley & Wasserman, 2008, p. 65

<sup>50</sup> See Mumford, 1998, p. 88 and Fara, 2005, p. 51

<sup>51</sup> Fara, 2005, p. 51

Suppose that some such list of conditions could be fully specified, and avoids the problem of triviality. Even if this much is granted, a further problem arises. The proposal would now be that fragility is correctly ascribed to a particular glass only if it would break when dropped in specific ‘ideal’ or ‘ordinary’ conditions: from a height of at least two metres above a surface at least as solid as normal concrete, through air of at least the normal density, not wrapped in packaging material, given the actual laws of nature, etc. But, granting these conditions, cases of masking – albeit of a different sort – are still possible. Suppose that the glass has a ‘strong spot’; when dropped in one precise way, when a particular part of it strikes the floor, at a particular angle, it would withstand much more force before breaking than it otherwise would, were it dropped in any other way, struck on any other part, at any other angle.<sup>52</sup> If so, a particular glass may be dropped in the specified ‘ideal’ conditions, and be disposed to break when dropped (in these conditions), but not break. This is a case of masking, not of finking, because the glass retains the intrinsic properties which are the causal basis of its disposition; the manifestation would be masked by the causal effectiveness of some other intrinsic property of that particular glass.

If these ‘ideal’ conditions are precisely those under which we would expect a fragile glass (of this sort) to break, and yet it does not break, can we really ascribe it with the disposition of fragility? After all, if this glass were dropped repeatedly in these precisely defined conditions (and it retained this property), it would not break. Note, however, that in these ‘strong spot’ cases, we are supposing that the glass would break in all other stimulus conditions within a precisely specified range (e.g. all other conditions are satisfied, but the angle of strike, say, is slightly altered). There is a reliable connection, but not a sure-fire connection, between the obtaining of this precise *range* of stimulus conditions, and the relevant manifestation. And, it seems that this is enough to grant that the glass is fragile, at least in a significant sense. We ought to deny it fragility only if we ascribe dispositions to objects by considering how they would react given *one* extraordinarily precisely defined scenario. In which sense do we ascribe dispositions?

There might be occasions on which we are interested in whether or not an object would have behaved in a certain way, given a single precisely defined scenario. But, this does not seem to be the case for all disposition ascriptions; in general, these ascriptions seem to be looser, permitting consideration of a range of stimulus conditions. Noting this, Whittle has suggested a helpful distinction between ‘global’ and ‘local’ dispositions.<sup>53</sup> An object has the global disposition of, say, fragility, if it has some property in virtue of which it would break in a suitable proportion of cases in which some stimulus conditions, within a specified range, obtain.<sup>54</sup> Global dispositions may be masked; objects may retain these dispositions consistent with the possibility of maskers which are not excluded in the specified range of conditions, and

---

<sup>52</sup> See Manley & Wasserman, 2008, pp. 69 – 70 (they call these ‘reverse Achilles heel’ cases).

<sup>53</sup> Whittle, 2010, p. 3

<sup>54</sup> This draws on Manley & Wasserman’s (2008) view; what counts as a ‘suitable proportion’ of cases depends upon the context of utterance (see p. 76). The global/local distinction need not be drawn on precisely these lines; see Whittle, 2010, p. 5.

with intrinsic maskers like strong and weak spots. By contrast, an object has the ‘all-in-local’ disposition of fragility (i.e. the canonical disposition ‘to-break-in-precise-conditions-C\*’) if it has some property in virtue of which it would break in the single, precisely specified, stimulus condition scenario.<sup>55</sup> All-in-local dispositions cannot be masked; if an object were to have a strong spot, such that it would not break in the precise specified scenario, it lacks this all-in-local disposition to do so. Hence, an object might have the global disposition of fragility without having the all-in-local disposition to break in specific conditions; this would be true in all cases in which a global disposition is masked. In some such cases, such as ‘strong spot’ cases, the object would have the all-in-local disposition to remain intact in these precisely specified conditions, insofar as it does so reliably in such conditions, in virtue of some intrinsic property; so, objects may have the all-in-local disposition to remain intact in such conditions, while having the global disposition of fragility.

Recognising this distinction enables us to agree that, insofar as we are talking about (non-probabilistic) all-in-local dispositions, then some conditional analysis which does not tolerate exceptions may analyse a disposition ascription. But, it seems that we are rarely talking about such dispositions; usually, when we attribute a disposition to an object, we are attributing a disposition which is not all-in-local, and which could be masked. Often we are attributing a global disposition, but there is a continuous spectrum of disposition-types, ranging from local to global, depending on the extent to which variation in stimulus conditions, within the specified range, is permitted.<sup>56</sup> Masking is possible in all but those cases of all-in-local dispositions.

For example, an object has a ‘fairly-local’ disposition of fragility if it has a property in virtue of which it would break in a suitable proportion of cases in which some stimulus conditions, within a specified range, obtain, and in which certain specified conditions are held fixed. Fairly-local dispositions are wider than all-in-local dispositions, as they permit variation of some conditions, within the specified range (i.e. those which are not held fixed). Hence, fairly-local dispositions can be masked; an object might have a fairly-local disposition to break in conditions C, but lack the all-in-local disposition to break in (more precisely specified) conditions C\*. But, fairly-local dispositions are narrower than global dispositions, as they include the stipulation that at least some conditions are held fixed – these conditions do not vary, even within the specified range. Hence, an object might have the global disposition of fragility without having the fairly-local disposition ‘to break in C’. If it were specified, in C, that the precise spot and angle of strike are fixed, and this ensured that the glass would always fall on its ‘strong spot’, then even if other conditions were varied within the specified range (e.g. the height of the drop, the quality of the floor

---

<sup>55</sup> This would need to be adapted in order to account for probabilistic dispositions. I have not discussed these as my primary interest in this chapter is with the kinds of dispositions to which compatibilists could appeal if determinism were true.

<sup>56</sup> Whittle, 2010, p. 8

surface, etc.) it may be the case that the glass would fail to break in a suitable proportion of these cases, and therefore lack the fairly-local disposition to break in C.

It seems, then, that dispositions – at least in a non-all-in-local sense – can be correctly ascribed to objects, despite there being no sure-fire connection between the obtaining of their specified stimulus conditions, and their manifestation.

### 5. Abilities, Conditionals, and Frankfurt-style Cases

Given the above discussion of disposition ascriptions, parallel issues may apply with regard to ability ascriptions. If so, this would be, at least *prima facie*, significant for Frankfurt's argument. Having noted the structural similarity between the Frankfurt-style cases and Martin's reverse-cycle electro-fink counterexample, there is an interesting divergence in the conclusions drawn; Martin's conclusion is that the simple conditional is an inadequate analysis of the disposition that the wire retains, while Frankfurt's conclusion is that the simple conditional is false and the agent did not retain the relevant ability. But, if an agent's ability to have acted otherwise can be finked or masked, such that the agent might retain it, in the relevant sense, consistent with its non-manifestation (i.e. their failure to act otherwise) given the stimulus conditions (i.e. when they try, etc.), then this Frankfurt-style case might not show what it is supposed to show; it might show, parallel with Martin's case, that the simple conditional is not an adequate analysis of the agent's ability, but not (or, at least, not obviously) that the agent lacked the relevant ability.

### 6. Finkish Abilities

Suppose that we maintain the parallel between Martin's reverse-cycle-fink case and this Frankfurt-style case; Dr. Yellow (or some device or mechanism of his) would, whenever Jones tries to do otherwise, cause some intrinsic change in Jones – alter, say, the state of affairs which obtains in Jones's brain – such that he would act in accordance with Dr. Yellow's plan. The simple conditional 'if Jones were to try to do otherwise, he would act otherwise' is therefore false. But, following Lewis's modified conditional analysis of dispositions, Vihvelin has argued in favour of the following modified conditional analysis of abilities: an agent has the ability, at some time  $t$ , to A if and only if the agent has some intrinsic property (or set of properties) B such that, if they were to try at  $t$  to A, and if they were to retain B, then B and the agent's trying at  $t$  to A would jointly cause the agent to A.<sup>57</sup>

On this view, when assessing whether an agent has the ability to A, one need consider whether they would A, when they try to A, in only those possible worlds in which (i) the actual laws of nature obtain,

---

<sup>57</sup> See Vihvelin, 2004, pp. 437 – 438

and (ii) the agent retains the relevant intrinsic properties. One may ignore worlds in which the agent loses these properties, as these are not worlds in which the relevant stimulus conditions obtain.<sup>58</sup>

It was granted that, in Dr. Yellow's absence, Jones could have acted otherwise than in accordance with Dr. Yellow's plan; on the proposed analysis, Jones had this ability in virtue of having some intrinsic properties which, together with his trying, would have caused his acting otherwise. Given Dr. Yellow's presence, Jones will not succeed in acting otherwise: if he tries to act otherwise, he will undergo intrinsic change, lose the relevant properties and thereby lose the ability to act otherwise; if he does not try to act otherwise, then he will not act otherwise. But, Dr. Yellow's *mere presence* makes only a counterfactual, not an actual, change to Jones's properties, and therefore what Jones is able to do. Dr. Yellow would become active and interfere with Jones's properties – thereby actually altering what Jones is able to do – only if Jones were to try to act otherwise. But, when Jones does not try to act otherwise, Dr. Yellow is dormant; Jones is intrinsically untouched by Dr. Yellow, and therefore retains whatever abilities he had in virtue of his intrinsic properties. The modified conditional seems, therefore, to be true of Jones, given Dr. Yellow's presence; he has properties, in Dr. Yellow's presence which, if he were to retain them, would, along with his trying to act otherwise, jointly cause him to act otherwise. And, at no point in the Frankfurt-style case does Dr. Yellow actually intervene; so, when Jones acted 'for reasons of his own' in accordance with Dr. Yellow's plan, he was able to have acted otherwise – Dr. Yellow's mere presence ensures only that Jones's ability was finkish and that Jones *would* not have acted otherwise, if he had tried.<sup>59</sup>

This analysis may, then, allow one to claim that Jones was able to have acted otherwise, at least in some sense, given Dr. Yellow's presence. In addition, it avoids a general objection, proposed by Keith Lehrer, against conditional analyses of abilities.<sup>60</sup> Lehrer argued that the proposition (i) 'S could have done A' cannot be analysed by any simple conditional of the form (ii) 'S would have done A, if some condition C had obtained' because (ii) is logically consistent with the conjunction of the following propositions:

(iii) 'S could have done A only if condition C had obtained'

(iv) 'Condition C did not obtain'

Propositions (ii) and (iii) are consistent because it is possible that the obtaining of condition C, which is (according to (ii)) sufficient for S to do A, is also necessary for S to be able to do A. And, the conjunction of (iii) and (iv) entails (v) 'S could not have done A', which is inconsistent with (i). So, the simple conditional analysis fails; (ii) does not entail (i).

To see the objection more clearly, consider the myth of the 'sword in the stone': a sword is stuck in a stone, and prospective kings are required to prove their worth by removing it. Suppose that Arthur is, in fact, extremely physically weak; Merlin thinks, nonetheless, that Arthur is the most suitable prospective

---

<sup>58</sup> See Smith, 2003, pp. 24 – 25

<sup>59</sup> Vihvelin, 2004, p. 438

<sup>60</sup> Lehrer, 1968, p. 43

king, and decides to ensure that Arthur will pass the required test, if he tries to do so. When, and only when, Arthur tries to pull out the sword, Merlin will endow him with the property of extraordinary physical strength, ensuring that he succeeds. Arthur's trying to pull out the sword, in these circumstances, is sufficient for his success. But, his trying is also necessary for him to be able to succeed; Merlin will endow him with this property only when he tries – otherwise, he is not strong enough to remove the sword. And, the simple conditional is true of Arthur at some time  $t$ , when he is not trying to pull the sword from the stone; this analysis, then, commits one to claiming that it is true, at time  $t$ , that Arthur can pull the sword from the stone, even if he is not, at time  $t$ , trying to do so. So, it seems that the truth of the simple conditional is insufficient for true ability ascriptions.

According to the modified conditional analysis, Arthur lacks the ability to pull the sword from the stone, but his lack of this ability is finkish. Arthur has no intrinsic property, the retention of which would, in combination with his trying, cause him to succeed. Merlin causes the necessary intrinsic change in Arthur when and only when Arthur tries to pull out the sword, and *this* new property, together with his trying, suffices for his success. Hence, Arthur acquires the ability to succeed only when he tries to do so; otherwise, he lacks the ability.

## 7. Masked Abilities

According to this Lewis-style analysis, an agent would lack the ability to do A if there were one possible world in which (i) the actual laws of nature obtain, (ii) the agent retains the relevant intrinsic properties B, (iii) conditions C obtain (i.e. the agent tries to A), and (iv) the agent does not do A. On this view, it is supposed that there is a sure-fire connection between the agent trying to do A, in the relevant possible worlds, and their doing A in these worlds.

The Frankfurt-style case may be re-described in such a way that Dr. Yellow's intervention would not cause Jones to lose the relevant intrinsic property B. Perhaps, on observation of Jones's decision or will to act otherwise, Dr. Yellow would physically overpower Jones, preventing him from acting otherwise than in accordance with the plan. Or, perhaps he might install some kind of force-field around Jones, such that no movement is permitted other than that which Dr. Yellow desires Jones to make.<sup>61</sup> This would be a scenario in which propositions (i)-(iv) are all true; the actual laws may obtain, Jones may try to act otherwise, and retain the relevant property B, but his trying, together with property B, would not cause him to act otherwise. On the Lewis-style view, then, Jones would lack the ability to act otherwise. Insofar as Jones is deemed responsible for what he has done in this case, then this version of Frankfurt's proposed counterexample may stand.

---

<sup>61</sup> Frankfurt's own description of the case does not contain enough detail to make it definitively a case of finking or a case of masking (see Clarke, 2009, p. 340). It suffices, for the present point, that the masking version is available, even if it is not Frankfurt's own version of the case.

Note, however, that this version of the Frankfurt-style case is analogous with the proposed cases of masking which were alleged to count against Lewis's analysis of dispositions. So, it might be apt to suggest that some abilities, like some dispositions, may be correctly ascribed to agents even though their manifestations would be masked in various possible scenarios, by factors which are consistent with their retaining the relevant intrinsic properties.<sup>62</sup> If so, then the truth of the Lewis-style analysis would be unnecessary for true ability ascriptions, of at least some sort; one may possess these abilities, despite there being no sure-fire connection between trying to do A, in the relevant possible worlds, and doing A.

Taking this view allows a modified response to this Frankfurt-style case. Fara has argued along the following lines: Jones retained, in the actual scenario, the ability to have acted otherwise, despite Dr. Yellow being present and ready to mask this ability. If the relevant stimulus conditions had obtained – had Jones tried to act otherwise, while retaining the relevant intrinsic properties – he would not have succeeded in acting otherwise, but this failure would not show that he lacked the ability to have acted otherwise; it would be a case in which his ability to act otherwise was masked by the activity of Dr. Yellow or his mechanism.<sup>63</sup>

## 8. General and Specific Abilities

There are, however, various senses in which we might claim that an agent was, or is, able to do A. Some of these senses seem to be compatible with masking; it is helpful to outline these senses, in order to be clear about the sense in which Fara's masking response entitles us to claim that Jones was able to have acted otherwise.

There is a familiar sense in which we ascribe an agent with the ability to ride a bicycle, even if we know that they currently lack access to a functioning bicycle and, if they were to try to ride a bicycle, they would fail to do so; similarly, we might say that one is able to walk, despite being tied to a chair, or able to speak French, despite being gagged.<sup>64</sup> We do so when we believe that an individual has acquired the relevant skill and are psychologically and physically suited to exercising it, when they try to do so, on some occasion (even if they do not (perhaps ever) do so). By analogy with the global dispositions discussed above, we might say that one actually possesses some such 'general ability' if one has some properties in virtue of which one would succeed in doing A in a suitable proportion of cases in which one tries to A, the actual laws of nature obtain, and one retains the relevant properties.<sup>65</sup>

---

<sup>62</sup> This is Fara's view – see his 2008, pp. 854 – 855.

<sup>63</sup> Fara, 2008, p. 855

<sup>64</sup> See van Inwagen, 1983, pp. 12 – 13

<sup>65</sup> We would not, for example, ascribe an individual with the ability, in this sense, to speak French on the basis of their doing so in some possible worlds in which they have acquired relevant skills that they lack in the actual world.

We might be more precise in specifying the relevant stimulus conditions. For example, we might ascribe a golfer with the ability to hole short putts, given the opportunity to do so (e.g. access to a putting green, club, ball, etc.). Again, such ascriptions are compatible with cases in which these conditions are satisfied, and in which the golfer fails to hole a short putt, when they try to do so. For example, a golfer may be such that they would hole a short putt in a large proportion of possible worlds in which (i) the actual laws obtain, (ii) the golfer retains the relevant properties, (iii) the golfer has the opportunity to hole a short putt, and (iv) the golfer tries to hole the putt. But, there are worlds within this set in which, for example, a gust of wind diverts the ball away from the hole, and the putt is missed. Intuitively, the golfer is nonetheless able to hole short putts, given the opportunity to do so. By analogy with more local dispositions, we might say that one actually possesses some ‘more specific’ (but, still quite general) ability if one has some properties in virtue of which one would succeed in doing A in a suitable proportion of cases in which one tries to A, the actual laws of nature obtain, one retains the relevant properties, and conditions are such that one has the opportunity to do A.

We might be even more precise in specifying the relevant stimulus conditions; one might ascribe an agent with the ability to do A, given the opportunity to do A, and given conditions x, y, and z. By specifying that conditions x, y, and z obtain, one need not have specified a complete set of conditions; specifying that conditions x, y, and z obtain in all worlds relevant to an assessment of whether or not the agent actually possesses this more specific ability is consistent with variation in other conditions. By analogy with fairly-local dispositions, we might say that one actually possesses some such ‘more specific’ ability if one would succeed in doing A, when one tries, in a suitable proportion of this narrow set of possible worlds. Insofar as some variation in conditions (aside from those specified as fixed) is permitted, it is not clear that any specified list of conditions could rule out all potential maskers.

When we ascribe an agent with an ability to do A, given a maximally precise specification of the relevant stimulus conditions, we ascribe an all-in ability to do A. We would ascribe an agent with, for example, the all-in ability to have acted otherwise, if we claimed that they were able to have acted otherwise, given conditions *precisely* as they were.

Abilities ascribed in all other senses (i.e. all but the all-in sense) specify a range of stimulus conditions, within which variation is permitted; hence, if determinism were true, different outcomes would nonetheless be possible, given the relevant stimulus conditions. This allows for the possibility of masking, even if determinism were true. But, if determinism were true, then the all-in ability to do A would be sure-fire – it could not be masked. If determinism were true, then the precise conditions which obtained at time  $t$ , together with the laws of nature, would have entailed precisely what occurred at some later time  $t_1$ . If the agent did not A at  $t_1$ , there is no nomologically possible world with conditions precisely as they were at  $t$  in which the agent does A at  $t_1$ . It follows from the nomological impossibility of one’s doing A,

given these precise conditions, that one lacks the all-in ability to do A. If it were nomologically possible that the agent does A at  $t_1$ , given the precise conditions at  $t$ , then it would be nomologically impossible that the agent does not A at  $t_1$ . So, insofar as one has the all-in ability to do A, one's ability is sure-fire – there are no nomologically possible worlds in which these precise conditions obtain at  $t$  and one tries and fails to A at  $t_1$ . This is not to say that, if determinism were true, one's trying to A would be sufficient for success; there might be cases in which one tries to A but fails due to some determined event. In all such cases one would lack the all-in ability to have done A; if determinism were true, then whenever one tried to do something but failed to do it, one lacked the all-in ability to do it.

According to Austin, however, all-in abilities need not be sure-fire; one may exercise an all-in ability in bringing about an outcome, without determining that the outcome occurs. Austin's golfer, who missed an easy putt, having tried to hole it, exclaims 'I could have holed it!'; in Austin's view, he attributes to himself the power to have holed that putt, given conditions precisely as they were.<sup>66</sup> This attribution implies, for reasons discussed above, that the miss was undetermined, given the precise conditions and the laws; if so, the golfer *might* have holed the putt, given conditions precisely as they were. But, even if we suppose that Jones has a maximally developed skill to hole putts of this sort, his trying would not guarantee success; at most, his trying might (to some degree of probability) issue in success. It might issue in failure, but such a failure would not always indicate that Jones lacked the all-in ability to hole the putt. Hence, if Austin is right about what is implied by attributions of ability, insofar as the golfer had the all-in ability to have holed the putt, this ability need not have been sure-fire.<sup>67</sup>

## 9. In What Sense Was Jones 'Able' to Have Acted Otherwise?

With these distinctions in mind, we can be clearer about the sense in which Fara's masking response permits one to maintain that Jones was able to have acted otherwise in the Frankfurt-style case.

Fara's attributions of masked abilities grant that the agent had the opportunity to exercise their ability, which is characterised as "the ability to do thus-and-so in certain circumstances"; one has the opportunity to exercise one's ability only if these circumstances obtain.<sup>68</sup> In one of his examples, an agent tries to unlock a door, with the correct key, but fails to unlock the door because an insect is crawling through the

---

<sup>66</sup> Austin 1956 (reprinted in his 1961), p. 218 (fn.).

<sup>67</sup> Granted, given conditions as they were, it was a matter of chance whether or not the putt was holed on this occasion. But, contra some (e.g. Berofsky, 2011, p. 156), we should not deny the golfer the all-in ability to have holed it. If he had holed it, we should grant that he was able to have holed it: he did so, and his trying was a cause of its being holed. We need not assume that causation requires determination (see Anscombe, 1971, pp. 87 – 104 (in Sosa & Tooley (eds.) 1993)). Success on a particular occasion does not entail an all-in ability to bring about success on that occasion; success might have been a fluke, an improbable possibility. But, we need not take Jones's success to be a fluke, even if – consistent with everything that he did – he might have failed. Just as we would not attribute an all-in ability on the basis of an isolated improbable success, we should not deny an all-in ability on the basis of an isolated improbable failure (see Maier, 2013, §1.3, and also Clarke, 2009, p. 346).

<sup>68</sup> Fara, 2008, p. 846

lock mechanism, preventing its normal functioning at the precise time of the agent's attempt.<sup>69</sup> The 'certain circumstances' here seem to be 'when one has access to the door, and the right key for its lock'; Fara suggests that an agent who has the wrong key would lack the ability, in the sense he is employing, to open the door – in his terms, they would lack the opportunity to do so, because the relevant circumstances do not obtain.<sup>70</sup>

This insect-induced failure to open the door does not seem to count against the agent's ability 'to open the door next to which one is standing, and for which one has the right key'.<sup>71</sup> There are many possible worlds in which these conditions obtain and no insect is crawling through the lock at the precise time of the agent's attempt. And, we may suppose that the agent would succeed in opening the door in a suitable proportion of the possible worlds in which these conditions obtain. So, the agent may actually possess this ability, despite the actual insect-induced failure.

Ought we to make a similar claim about Jones's ability to have acted otherwise in the Frankfurt-style case? If the specified circumstances include the presence of Dr. Yellow, then it seems that we ought not to. In all relevant possible worlds (i.e. those in which Jones retains the relevant properties, and the actual laws obtain) in which the facts about Dr. Yellow are as specified in the original story, Jones would not succeed in completing an alternative type-action, or even in making an overt attempt to do so, were he to try to do so – he would not act otherwise. So, insofar as we are considering the fairly specific ability to have acted otherwise *given the presence of Dr. Yellow* (as specified in the original story), we should conclude that Jones lacked this ability in the Frankfurt-style case; Dr. Yellow was not merely a potential masker of this ability, but in fact removed it.

This is consistent with claiming that Dr. Yellow was merely a potential masker of an ability to have acted otherwise – in a more general sense – which Jones retained in the Frankfurt-style case. For example, if Dr. Yellow's presence is not specified in the stimulus conditions, then a wider set of possible worlds may be considered – for example, those in which the actual laws obtain, Jones tries to act otherwise, and retains the relevant set of properties. And, we may suppose that Jones would succeed in acting otherwise in a suitable proportion of these worlds. His failure to act otherwise, when he tries to do so, in worlds which contain Dr. Yellow (as specified in the original story), even if one such world was actual, would not count against Jones's continued possession of this ability; his ability would merely have been masked by Dr. Yellow's activity in these cases. Evidently, then, Fara is best interpreted as making the claim that Jones retains, in the Frankfurt-style case, a general ability to have acted otherwise, which might be masked by Dr. Yellow.<sup>72</sup>

---

<sup>69</sup> Ibid. p. 847

<sup>70</sup> Ibid. p. 847

<sup>71</sup> It would seem to count against the agent's having the all-in ability to have opened the door.

<sup>72</sup> See Whittle, 2010, p. 9

Given the masking objection against the Lewis-style analysis of abilities, an alternative response is available; one might attempt a defence of the Lewis-style analysis, analogous with Lewis's own proposed defence of his analysis of dispositions – correctly specified stimulus conditions ensure a sure-fire connection between the occurrence of the stimulus conditions and the manifestation of the ability. In this version of the Frankfurt-style case (in which Dr. Yellow is considered as a potential masker), Jones would fail to act otherwise (even when he tries to do so and retains all relevant properties) in all nomologically possible worlds in which Dr. Yellow is present (and as he was specified in the original story). So, for the Lewis-style analysis to hold, it must be the case that, when we ascribe Jones with the ability to act otherwise, we are specifying, in the relevant stimulus conditions that, amongst other things, Dr. Yellow is absent or relevantly different.

In attempting to defend the Lewis-style analysis it becomes clear that it cannot support the view that Jones retained, in the actual scenario, the ability to have acted otherwise given the presence of Dr. Yellow. And, therefore, one must concede that Jones was not able to have acted otherwise in the actual scenario, as Dr. Yellow was actually present. At most, one can maintain that Jones retained, in the actual scenario, an ability to have acted otherwise in a specified set of conditions which did not actually obtain – conditions which include the absence of Dr. Yellow.

There are, of course, nomologically possible worlds in which Dr. Yellow is present, but does not actually prevent Jones from acting otherwise. Some of these worlds are consistent with the facts of the Frankfurt-style case; for example, worlds in which Jones does not try to do otherwise (i.e. in which he tries to, and does, act in accordance with Dr. Yellow's plan). Others are worlds in which, inconsistent with the facts of the Frankfurt-style case, Dr. Yellow is not concentrating, or in which his mechanism malfunctions; in these worlds, Dr. Yellow would not prevent Jones from acting otherwise when he (Jones) tries to do so.

Given that the latter set of scenarios is possible, one should agree that Jones did retain, in the actual scenario, an ability to have acted otherwise given Dr. Yellow's presence *and inactivity*. Some have taken the fact that, in the actual scenario, Dr. Yellow was inactive (he did not actually intervene), to indicate that Jones did indeed retain, in this scenario, an ability to have acted otherwise *in this scenario*.<sup>73</sup> But, this is false. Granted, Dr. Yellow was inactive in the actual scenario; this was because the stimulus for his intervention – Jones trying to do otherwise – did not actually obtain. The facts of the Frankfurt-style case specify that Dr. Yellow would be inactive *only* when this stimulus does not obtain. So, when assessing what Jones is able to do, in the actual scenario, the observation of Dr. Yellow's actual inactivity licences consideration of only those possible worlds in which the stimulus does not obtain (i.e. in which Jones does not try to do otherwise) – not worlds in which the stimulus *does* obtain and Dr. Yellow is nonetheless inactive (e.g.

---

<sup>73</sup> See, for example, Campbell, 1997, pp. 325 – 326, and Smith, 1997, pp. 311 – 312

worlds in which, inconsistent with what is specified in the Frankfurt-style case, Dr. Yellow is lazy, or his mechanism malfunctions). By considering what would occur in this wider set of worlds, one is supposing that some of the conditions specified in the Frankfurt-style case do not obtain; one would, therefore, be claiming that Jones retained, in the actual scenario, an ability to have acted otherwise in conditions which could not obtain, consistent with what is specified in the case. Jones did not retain, in the actual scenario, an ability to have acted otherwise, given Dr. Yellow's presence, *and* given all of the facts about Dr. Yellow, specified in the original story.

#### 10. Is This Sort of Ability Relevant to PAP?

On neither strategy, then, is one entitled to claim that Jones retained, in the Frankfurt-style case, the ability to have acted otherwise in Dr. Yellow's presence (given the facts about Dr. Yellow specified in the original story). We should grant that Jones retained, in the actual scenario, the ability to have acted otherwise, in a general sense; he retained this ability in virtue of being such that, in a suitable proportion of nomologically possible worlds in which he retains the relevant properties and in which he tries to act otherwise, he would act otherwise.

The respective strategies of Vihvelin and Fara are claimed to vindicate PAP against Frankfurt's challenge; the ability to have acted otherwise, which Jones retained, is said to satisfy the consequent of PAP, and thus to show that Frankfurt's challenge fails to count against the leeway condition for responsibility.<sup>74</sup> This amounts to a defence of the following version of PAP: one is responsible for what one has done only if one could (was able to) have acted otherwise, in a general sense (i.e. one would have acted otherwise, had one tried to do so, in a suitable proportion of possible worlds in which the actual laws obtain and one retains the relevant properties). Call this 'the general version' of PAP.

Frankfurt might, however, have been happy to concede that this version of PAP is not falsified by the proposed case. Nonetheless, advocates of these strategies should concede that a different version of PAP *is* falsified by the Frankfurt-style case – PAP interpreted as follows: one is responsible for what one has done only if one could (was able to) have acted otherwise, in a more specific sense (i.e. one would have acted otherwise, had one tried to do so, in a suitable proportion of possible worlds in which certain other conditions (beyond the actual laws and the agent's relevant properties) obtain; in this case, the relevant facts about Dr. Yellow). If this version of PAP was the target of Frankfurt's challenge, then the kind of responses under consideration do not show that the challenge fails.<sup>75</sup>

---

<sup>74</sup> Vihvelin, 2004, pp. 428 – 429, Fara, 2008, p. 855 (see also Smith, 2003, p. 19).

<sup>75</sup> See, for example, McKenna, 1998, pp. 261 – 262, Clarke, 2009, p. 341, and Fischer, 2011, p. 262

The target of Frankfurt's challenge was leeway theory in general; he granted that the compatibility of determinism and the relevant ability to have done otherwise was a matter of dispute, and argued for the broad claim that one may be responsible for what one has done even if one could not have done otherwise.<sup>76</sup> Insofar as one takes the truth of determinism to threaten anyone's possession of the relevant ability, one must endorse a specific-ability interpretation of PAP; one might also endorse the general version, but it seems clear that the truth of determinism would not threaten agents' possession of a general ability to have acted otherwise. Leeway incompatibilists must claim that the 'all-in' ability to have acted otherwise occupies the consequent of PAP. The fact that the proposed counterexample does not show that Jones lacked a general ability to have acted otherwise would not, then, count against the conclusion (if it were warranted by the Frankfurt-style case) that the all-in ability to have acted otherwise is unnecessary for responsibility, and that this version of PAP is, therefore, false.<sup>77</sup>

Vihvelin and Fara both identify themselves as leeway compatibilists.<sup>78</sup> They must not, therefore, claim that the consequent of PAP is occupied by the 'all-in' ability to have acted otherwise, as possession of this ability is incompatible with determinism. The kind of ability which occupies the consequent of a compatibilist's version of PAP must be less-specific; it must allow at least some variation within a range of specified stimulus conditions. It does not follow, however, that leeway compatibilists can endorse only the general version of PAP, and no more specific version. It is consistent with leeway compatibilism to consider the agent to be responsible for what they have done only if they could have acted otherwise, given *some* conditions held fixed; hence, the ability to have acted otherwise, given Dr. Yellow's presence, may be required in order satisfy PAP in the Frankfurt case, consistent with leeway compatibilism.<sup>79</sup> The fact that the proposed counterexample does not show that Jones lacked a general ability to have acted otherwise would not count against the conclusion that the more specific ability to have acted otherwise (i.e. given Dr. Yellow's presence) is unnecessary for responsibility, and that this more specific (but nonetheless compatibilist) version of PAP is, therefore, false.

This establishes that leeway compatibilists may endorse versions of PAP which are more specific than the general version. But, it does not establish that there are no independently plausible versions of leeway compatibilism which accept only the general version of PAP. If so, these views might stand against Frankfurt's challenge.

There are reasons to doubt that such views are independently plausible. Consider a case of actual intervention.<sup>80</sup> Suppose that an agent – Ben – has the general ability to pick up a red sweet when presented with a choice between red and blue sweets; in a suitable proportion of relevant possible worlds

---

<sup>76</sup> Frankfurt, 1969, p. 1 (see also Vihvelin, 2004, p. 446)

<sup>77</sup> I consider, in the third chapter, whether Frankfurt-style cases do warrant this conclusion.

<sup>78</sup> Vihvelin, 2008, p. 343, Fara, 2008, pp. 860 – 863

<sup>79</sup> Whittle, 2010, p. 10

<sup>80</sup> See Whittle's 'Bound Ben' example – Ibid. p. 10 – hers is a case of responsibility for an omission.

in which he tries to do so, he would succeed in doing so. But, in actual fact, when Ben tries to pick up a red sweet, he is physically compelled, via the manipulation of some other agent, to pick up a blue sweet. A leeway compatibilist who endorses *only* the general version of PAP as the freedom-relevant condition for responsibility must claim that Ben satisfied this condition for having picked up a blue sweet; he retained, in the actual scenario, a general ability to have acted otherwise.

But, this seems to be highly counterintuitive; intuitively, Ben does *not* satisfy the freedom relevant condition in the actual (manipulation) scenario, and would do so in a counterfactual scenario in which he picked up a blue sweet without being manipulated. This leeway compatibilist cannot account for this apparent distinction because, in both cases, Ben would have retained a general ability to have acted otherwise, and thereby satisfied *their* freedom relevant condition for responsibility. In order to account for the distinction, one would need to maintain that some ability, the possession of which would be undermined by actual manipulation, is necessary for Ben's to be responsible for what he has done; and, this would be to endorse some more specific version of PAP.

One might resist this conclusion by claiming that, in the actual scenario, Ben did not retain the general ability to have picked up a red sweet. If this were the case, he would fail to satisfy the general version of PAP and this leeway compatibilist could thereby account for the intuitive distinction between cases. But, it is unclear, given much of the preceding discussion, why we should consider manipulation as a temporary remover, rather than merely a masker, of Ben's general ability.<sup>81</sup> We need not suppose that the manipulation alters Ben intrinsically.

Of course, the intervention in the Frankfurt-style case is merely counterfactual, whereas in this example it is actual. But, the example is merely employed in an attempt to illustrate that leeway compatibilists ought to consider some fairly specific ability to be necessary for responsibility; the most independently plausible leeway compatibilist views are those which endorse a more specific version of PAP. If so, we may consider Frankfurt's challenge to target the strongest leeway compatibilist views, and the observation that Jones retained a general ability to have acted otherwise in the Frankfurt-style case poses no immediate threat to Frankfurt's rejection of the necessity for responsibility of the agent's being able to have acted otherwise *in some more specific sense*.

#### 11. Significance of General Abilities for the Actual Sequence

The observation that Jones, in the Frankfurt-style case, did retain a general ability to have acted otherwise does not seem to vindicate a plausible version of PAP against Frankfurt's challenge. It may, however, be significant in accounting for the intuition that Jones is responsible, in this case, for what he did. Rather

---

<sup>81</sup> Ibid. p. 11

than indicating that there were genuine practical alternatives accessible to Jones at the time of his action, it may indicate something about the kind of power which the agent actually exercised when acting as he did.<sup>82</sup>

We may suppose that Jones was psychologically such that he possessed a normally developed set of general human abilities; amongst this set were general abilities to recognise factors as counting in favour of a certain course of action, and to act on the basis of one's recognition of these factors.<sup>83</sup> He was not, in P.F. Strawson's terms, 'merely a child' or 'a hopeless schizophrenic', so our attitudes towards him, in response to his behaviour, were not modified on this basis – he was not exempt in general from certain demands on his conduct.<sup>84</sup> But, as is clear from the case of actual intervention, retaining such general abilities is not sufficient for satisfaction of the freedom-relevant condition on responsibility. Consistent with this, one's actual exercise of one's general abilities – to recognise reasons and to act accordingly – might have been impaired; for example, if one was placed under great strain, or forced to act in a certain way.<sup>85</sup> Our attitudes, in response to one's behaviour, may be modified by our recognition of such circumstances.

In the actual intervention case, one's attempts are *actually* pre-empted or frustrated, and some other mechanism is actually effective in bringing about one's action. But, in the Frankfurt-style case, Jones's exercise of his general abilities is, in actual fact, uninhibited; he actually sees most reason to act as (unbeknownst to him) Dr. Yellow wants him to act, and does so. Had he seen most reason to act in any other way, then Dr. Yellow would have actually intervened, and Jones would not have been responsible for what he would have been caused, by Dr. Yellow, to do. But, in actual fact, he brought about his action via his exercise of an ability which he would have successfully exercised in various different ways, had conditions been different – that is, had he seen most reason to act differently, tried to act differently, and had Dr. Yellow been absent. Jones's retention of this ability is precisely what Vihvelin and Fara are highlighting.

Yet, those who reject PAP, and advocate (along with Frankfurt) some version of an 'actual-sequence' theory of responsibility, have made this point in support of *their* view. On Fischer and Ravizza's view, for example, the psychological processes which are actually effective in bringing about one's action are called the 'mechanism' that leads to one's action.<sup>86</sup> They claim that the freedom-relevant condition on responsibility is satisfied when one's action is brought about via one's own moderately reason-responsive mechanism.<sup>87</sup> The 'reason-responsiveness' of a mechanism is to be understood in terms of the

---

<sup>82</sup> McKenna, 1998, p. 260

<sup>83</sup> These correspond to the capacities Fischer & Ravizza call 'receptivity' and 'reactivity' – see their 1998, pp. 69 – 76

<sup>84</sup> See Strawson, 1962, (reprinted in Watson, 1982), p. 67

<sup>85</sup> Ibid. p. 67 (see also, Campbell, 2005, pp. 401 – 402).

<sup>86</sup> Fischer & Ravizza, 1998, p. 38

<sup>87</sup> Ibid. pp. 45 – 46 (also pp. 81 – 82)

dispositional properties of that mechanism.<sup>88</sup> These properties are assessed by asking how the mechanism (with its relevant properties held fixed) would respond in various possible worlds, in which various different considerations might be taken to count in favour of different courses of action. Moderate reason-responsiveness requires that there is suitable set of worlds in which the mechanism recognises when there is reason to act otherwise, and that the agent would act otherwise in at least one of these worlds.<sup>89</sup>

The dispositions in question are general, such that the possible worlds deemed relevant to their assessment are those in which the mechanism operates unimpeded; worlds in which a Frankfurt-style intervener is present are not included.<sup>90</sup> Fischer and Ravizza agree that, in a Frankfurt-style case, the agent acted via a mechanism which was responsive to reasons, in the right way; the mechanism would have, in a suitable set of possible worlds, recognised reasons to act otherwise, and would have brought about an alternative action in at least one of these worlds. But, it does not follow that the agent could have acted otherwise, in any relevant sense; alternative scenarios in which the mechanism does bring about some alternative action were not ‘genuinely accessible’ to the agent – they did not “have it in their power to actualise such scenarios”, because had they seen more reason to act otherwise, the operation of the mechanism would have been impeded by Dr. Yellow.<sup>91</sup>

It seems, then, that the style of argument which I have been considering may be employed by those who reject PAP, in order to highlight properties of the mechanism via which the agent acted in the actual scenario. Its prospects as a vindication of PAP are, therefore, poor.

## 12. The Ability to Try

So far, I have been considering analyses of the ability to successfully perform an alternative type-action, or at least to make an overt attempt to do so; I have claimed that these analyses fail to show that agents in Frankfurt-style cases retain an ability of the relevant sort. There may be further reason to find these analyses to be inadequate; the truth of the analysans (that the agent would succeed in acting, or at least beginning to act otherwise, if they were to try to do so, and their ability were not finked or masked) may be insufficient for the truth of the analysandum (that the agent is able to act otherwise) because it is consistent with the agent being unable to try to act otherwise.<sup>92</sup>

---

<sup>88</sup> Ibid. p. 53

<sup>89</sup> Ibid. pp. 70 – 73

<sup>90</sup> McKenna, 2011, p. 190

<sup>91</sup> Ibid. p. 53 (see also Hurley, 2000, pp. 205 – 206).

<sup>92</sup> Chisholm, 1964 (reprinted in Watson, 2003), p. 29 – ‘try’ may be replaced by ‘decide’ or ‘choose’ (see also, Lehrer, 1968 (reprinted in Watson, 1982), p. 44 & van Inwagen, 1983, p. 119).

For example, suppose that Alice has a severe phobia of spiders; she is psychologically such that she could not (in some sense) try to pick up a spider.<sup>93</sup> It seems that Alice lacks the ability to pick up a spider. But, if she were to try to pick up a spider, in certain circumstances, she would succeed (in the absence of finks or maskers) – she has the required bodily properties and skills, and her trying would indicate the absence of her phobia. Hence, the proposed analysis seems to commit its adherents to the view that Alice, with her phobia, *has* the ability to pick up a spider; yet, surely she lacks this ability – this seems like a counterexample to sufficiency of the proposed analysis.

Fara concedes that, in this case, Alice lacks the ability to pick up a spider, but argues that his analysis of the ability to pick up a spider does not commit him to the view that Alice has this ability. Given Alice's phobia, the stimulus conditions for this ability, which include Alice's trying to pick up a spider, could not (in some sense) occur; on Fara's view, objects lack dispositions, and (therefore) agents lack abilities, if the relevant stimulus conditions are impossible.<sup>94</sup> There are various senses in which an occurrence may be said to be impossible; Fara agrees that it is both metaphysically and nomologically possible for Alice to try to pick up a spider.<sup>95</sup> He adds that it is 'notoriously difficult' to be clear about the sense in which it is impossible for Alice to try to pick up a spider, and claims, somewhat obscurely, that there is *some* non-trivial sense in which it is impossible.

This response seems to be problematic, both in its own terms and with regard to Fara's proposed treatment of Frankfurt-style cases. Firstly, it is not obvious that objects do lack relevant dispositions when it is impossible, but not metaphysically or nomologically impossible, for their stimulus conditions to occur. Fara gives the example of a rubber ball nailed to a wall, which cannot (in the relevant sense) be dropped, and claims that it lacks the disposition to bounce when dropped.<sup>96</sup> But, presumably the ball is disposed to bounce *if it were dropped*, in specified conditions; we may suppose that it is intrinsically such that it would do so, in a suitable proportion of nomologically possible worlds.<sup>97</sup> Perhaps Fara is denying that the ball has the all-in-local disposition of 'bounciness'; perhaps this is correct, but there are various less-local dispositions which the ball seems to retain. And, these are the sorts of dispositions which Fara has suggested constitute the relevant ability to have acted otherwise in the Frankfurt-style case.<sup>98</sup> So, presumably he should also agree that, on his analysis, Alice retains the ability (in some broader sense) to pick up a spider, despite lacking the ability to try to do so. This need not commit Fara to a counterintuitive claim about Alice's abilities; perhaps those who press the sufficiency objection concede that Alice may retain an ability of this sort. But, as discussed, this is not obviously the kind of ability relevant to freedom and responsibility.

---

<sup>93</sup> See Fara, 2008, p. 851

<sup>94</sup> Ibid. pp. 851 – 852

<sup>95</sup> Ibid. p. 852

<sup>96</sup> Ibid. p. 852

<sup>97</sup> See Clarke, 2009, pp. 335 – 336

<sup>98</sup> Fara, 2008, p. 855

Secondly, there are Frankfurt-style cases which purport to show that the agent could not have decided or chosen otherwise than they did; the third Frankfurt-style case presented in the first chapter – the case of Jones and Dr. Black – is one such case. Deciding to do A and choosing to do A are not obviously to be identified with one another, or with trying to do A.<sup>99</sup> But, such cases may certainly be adapted, such that Dr. Black's presence purports to ensure that Jones could not (in the relevant sense) have tried to act otherwise. Fara has acknowledged that there are such Frankfurt-style cases – in which it is not the case that Jones would try to act otherwise, and fail – and has claimed that his arguments “apply equally to these variant cases”.<sup>100</sup> But, this claim is not obviously consistent with his proposed response to the case of Alice. In these Frankfurt-style cases, there is also a sense in which the relevant stimulus conditions (which include Jones trying to act otherwise) are impossible (but metaphysically and nomologically possible): if Jones were *about to* try to act otherwise, Dr. Black would intervene and make Jones try to act in accordance with his plan. If Fara wishes to maintain that Jones retains the ability to act otherwise in these cases, and that this ability is potentially masked by Dr. Black, then what distinguishes these cases from that of Alice, in which Fara wishes to maintain that she lacks the ability to pick up a spider?

Perhaps there is a relevant disanalogy between these cases, such that Fara's claims about Alice's lack of the ability to try to pick up a spider do not commit him to a parallel conclusion about Jones in the Frankfurt-style case.<sup>101</sup> Even so, unless he wishes to argue that one might retain the ability to act otherwise without being able to try to act otherwise, Fara must claim (insofar as he wants to maintain his view about Frankfurt-style cases) that Jones retained the ability to have tried to act otherwise, in the presence of Dr. Black. But, it is not obvious that he is entitled to make this claim. This point has some added significance; as mentioned in the first chapter, if Jones were to have retained the ability, in the Frankfurt-style case, to have tried to act otherwise (or to have decided, or chosen, to act otherwise) this, in itself, might satisfy a plausible version of PAP. Trying, deciding, and choosing may be regarded as mental actions, and the ability to have tried, decided, or chosen to act otherwise may, therefore, be regarded as an ability to do otherwise, relevant to one's responsibility for what one has done.<sup>102</sup>

Given that Jones will not try to act otherwise in Dr. Black's presence (because, if he were about to do so, Dr. Black would intervene and prevent him from doing so) presumably Fara must claim that Jones's ability is masked, not removed, by Dr. Black. But, Fara has not attempted to analyse the ability to try to perform an action; he states that his proposed analysis is limited to abilities to perform actions “that one can in principle try to perform”, and acknowledges that one cannot in principle try to try to perform an

---

<sup>99</sup> See Clarke, 2009, p. 336 (fn.)

<sup>100</sup> Fara, 2008, p. 854 (fn. 13)

<sup>101</sup> See Whittle, 2010, pp. 16 – 18 for a discussion of this; Whittle claims that there is no such disanalogy.

<sup>102</sup> See Clarke, 2009, p. 347 and Mele, 2003, c. 9

action.<sup>103</sup> And, given that his analysis of masking includes the specification that the agent tries (and fails) to perform the action in question, it seems clear that an agent's ability to try cannot be masked in the way that Fara describes.

This leaves it open that this, or some other relevant ability, might be analysed in some other way, with stimulus conditions which do not include trying. Vihvelin, for example, has suggested that the relevant ability is 'to make choices on the basis of reasons'; on her view, to have this ability is to have a bundle of dispositions, including (i) to form and revise beliefs on the basis of evidence and argument, and (ii) to choose (or try) to act in response to one's pro-attitudes and beliefs.<sup>104</sup> If it is claimed that one has the ability to choose to act otherwise if and only if one would choose to act otherwise in a suitable proportion of nomologically possible worlds in which one retains the relevant properties and has specified pro-attitudes and beliefs, then one's ability to choose to act otherwise may be maskable. For example, one may fail to choose to act otherwise, given these conditions, in worlds in which Dr. Black is present and the relevant facts about him obtain; one may, nonetheless, retain the ability to choose to act otherwise, because one would succeed in a suitable proportion of relevant worlds.

This would entitle one to claim that Jones was able, in a general sense, to have chosen to act otherwise in the Frankfurt-style case. But, as discussed above, this is not obviously the relevant kind of ability; when we consider practical alternatives to be relevant to assessing Jones's responsibility for what he has done, we seem to be interested in whether or not it was within Jones's power to have done otherwise (including to have tried, decided, or chosen otherwise) given the specified conditions. And, this style of argument does not show that Frankfurt's claims, regarding the agent's lack of *this* kind of ability in his examples, are false.

---

<sup>103</sup> Fara, 2008, p. 849 (see Clarke, 2009, p. 347 for agreement on this point, and for analogous claims about deciding). If one thinks that one can, in principle, try to try to perform an action, then one would be pressed to show that one retains this ability in problem cases like Alice's – these cases can be formulated such that a regress develops.

<sup>104</sup> Vihvelin, 2004, pp. 439 – 440

## Chapter 3 – Leeway Incompatibilism and the Power to Determine One’s Will

### 1. Two Kinds of Intervention

Some leeway compatibilists have denied that an agent’s trying, deciding, or choosing to act constitutes a voluntary exercise of an ability of that agent.<sup>105</sup> They have considered the agential ability, relevant to attributions of freedom and responsibility, to be at the level of overt actions, rather than at the level of the will.<sup>106</sup> Frankfurt-style cases in which the intervener is, like Dr. Yellow, capable only of intervening after the agent has tried (etc.) to act otherwise, challenge leeway compatibilist views of this sort. I have argued that, for all that has been said in their defence by those who draw strong parallels between dispositions and abilities, Frankfurt’s challenge succeeds against such views. But, such views are not compulsory for leeway compatibilists; one may consider an agent’s trying (etc.) to perform an action to be an exercise of that agent’s ability to try to perform this action, and one may consider this to be a kind of ability relevant to the satisfaction of PAP. As discussed, leeway compatibilists might provide a distinct analysis of the ability to try – or, more commonly, the ability to choose on the basis of reasons – and claim that agents may have the ability to choose to act otherwise, consistent with the truth of determinism.<sup>107</sup> Leeway incompatibilists have characteristically considered an agent’s ability to try, decide, or choose (i.e. to exercise their *will*) to act otherwise to be relevant to an agent’s freedom and their satisfaction of PAP; they locate the agential ability, relevant to freedom and responsibility, at the level of the will, rather than only at the level of overt actions.<sup>108</sup>

Dr. Yellow’s presence renders true the following conditional: (i) for every overt action contrary to the intervener’s plan, if the agent were to try, decide, or choose (etc.) to perform this action, the intervener would prevent the agent from successfully performing it, and instead ensure that they successfully act in accordance with the plan.<sup>109</sup> The trigger for Dr. Yellow’s intervention is the fact that Jones has begun an attempt to act otherwise. Insofar as Dr. Yellow is capable of this and only this sort of intervention, and insofar as we understand Jones’s trying, deciding, or choosing to act otherwise to be an exercise of his *will* to act otherwise, Dr. Yellow must fail to ensure that Jones could not have done otherwise in *any* sense relevant to judgments about his responsibility for what he has done. In the actual scenario, Jones decided

---

<sup>105</sup> Hobbes argued that the ‘will’ is nothing more than a passive appetitive capacity of agents. See Hobbes, *Of Liberty and Necessity*, §27 (Chappell (ed.), 1999 p. 37) – I go on to discuss some of his arguments in §11.

<sup>106</sup> As discussed (see fn. 36), commitment to empiricism has motivated scepticism about irreducible potentialities of objects; this scepticism extended to agents’ wills, leading to focus on freedom of action as the relevant agential ability - simple conditional analyses of the ability to *act* otherwise promised to ensure a significant sort of freedom, without entering into ‘obscure and intractable’ debate about the ‘will’ (see Vihvelin, 2004, p. 441).

<sup>107</sup> See, for example, Wolf, 1990, p. 115, and Vihvelin, 2004, pp. 443 – 444

<sup>108</sup> See, for example, Clarke, 2009, p. 347

<sup>109</sup> Vihvelin, 2000, pp. 9 – 10. She calls this ‘conditional intervention’.

to eat a blue sweet, having lacked the ability to succeed in acting otherwise (in the specified conditions), but having retained the ability to *exercise his will* to act otherwise.

There are Frankfurt-style cases which employ more sophisticated interveners – for example, the case of Jones and Dr. Black – which purport to show that, given the intervener’s presence, the agent could not have ‘willed’ to act otherwise, and is nonetheless responsible for what they have done.<sup>110</sup> In the case that I presented, Dr. Black’s presence was supposed to render true the following conditional: (ii) for every action (including ‘acts of will’) contrary to the intervener’s plan, if the agent would otherwise (i.e. in the absence of intervention) go on to perform this action, the intervener would prevent the agent from doing so, and instead make them act in accordance with the plan.<sup>111</sup>

The trigger for Dr. Black’s intervention is the purported fact (or, perhaps, just Dr. Black’s belief) that Jones would otherwise go on to act contrary to his plan, in some relevant sense.<sup>112</sup> This, in theory, allows Dr. Black to pre-empt any relevant activity of Jones’s will, and to rule out the kind of apparently freedom-relevant ability to do otherwise that Jones may have retained in the Dr. Yellow case. There is no reason to suppose that Dr. Black might not make both (ii) and (i) true. But, if he does make (ii) true, then he need not make (i) true in order to be an effective intervener; given (ii), Jones would not even begin to act otherwise. And, if he does not make (ii) true, then it seems that his making (i) true would be of little significance, as it would leave intact Jones’s ability to at least begin, if not to succeed in, trying, deciding (etc.) to act otherwise. Of primary significance, then, is the matter of whether or not a Frankfurt-style case can be specified such that a Dr. Black-type intervener does make (ii) true.

## 2. The Power to Predict and Pre-empt Activity of the Agent’s Will

It is reasonably clear how Dr. Yellow might ensure that Jones will not succeed in performing any alternative type-action (i.e. denying him what Frankfurt has called ‘freedom of action’), while allowing Jones to ‘act freely’, in accordance with his (Dr. Yellow’s) plan; Jones would not act otherwise without initially trying (etc.) to do so, and Dr. Yellow would intervene immediately when, and only when, Jones tries (etc.) to act otherwise.<sup>113</sup> It is not so clear how Dr. Black might ensure that Jones will not exercise his will to act otherwise (i.e. denying him what Frankfurt has called ‘freedom of the will’), whilst allowing

---

<sup>110</sup> Or even begun to will to act otherwise, if this is distinguishable from succeeding in doing so.

<sup>111</sup> Ibid. pp. 9 – 10. She calls this ‘counterfactual intervention’.

<sup>112</sup> It may be that statements about the future lack, at the time of utterance, a determinate truth-value. Hence, it might not be a fact, at  $t$ , that Jones would or would not otherwise go on to act in a certain way at some later time  $t_1$  (see Torre, 2011, for a discussion of relevant issues).

<sup>113</sup> On Frankfurt’s usage, one ‘acts freely’ when one is unconstrained to act as one wills; one has ‘freedom of action’ only if, in addition, one could have acted otherwise. On a Hobbesian view, acting freely is all that is required for ‘freedom’.

Jones to ‘act of his own free will’, in accordance with his (Dr. Black’s) plan.<sup>114</sup> The power to observe and react to activity of Jones’s will is insufficient; intervention based on the observation that Jones has initiated a decision to act otherwise would be too late to ensure that Jones could not have exercised his will to act otherwise, in some relevant sense. In order to deny Jones ‘freedom of the will’, Dr. Black must have the power to predict and pre-empt the relevant activity of Jones’s will.

Frankfurt was initially imprecise on how Dr. Black is supposed to achieve this; he stated merely that “Black is an excellent judge of such things”.<sup>115</sup> In a footnote he elaborated: we are to suppose that Dr. Black has observed Jones deliberating over these options many times previously, and on every occasion on which Jones decided, at time  $t_2$ , to eat a blue sweet, he twitched involuntarily at a prior time  $t_1$ ; there were no occasions on which Jones decided at  $t_2$  to eat a red sweet, or no sweets, and twitched at  $t_1$ .<sup>116</sup> Dr. Black uses the twitch (or some other ‘prior-sign’) as a basis for prediction – he will intervene and determine Jones’s will when, and only when, he observes no twitch at  $t_1$ .<sup>117</sup> Nothing in the description so far shows that a scenario in which Jones twitches at  $t_1$ , and goes on to decide otherwise at  $t_2$  is, in any sense, impossible. Granted, Dr. Black has observed a correlation between these events, but we have no reason to suppose that not twitching at  $t_1$  is necessary for Jones to decide otherwise at  $t_2$ , and therefore no reason to suppose that Jones could not have decided otherwise at  $t_2$ , having twitched at  $t_1$ .

Frankfurt elaborated further by supposing that “there is some causal relation between Jones’s state at the time of the twitch and his subsequent states”.<sup>118</sup> If this causal relation were deterministic, then Dr. Black’s predictions would be reliable; for example, it would be nomologically impossible for Jones to decide otherwise at  $t_2$ , given that he has twitched at  $t_1$ , if the occurrence of any state of affairs at  $t_1$  which includes Jones twitching, along with the actual laws of nature, entails that a state of affairs obtains at  $t_2$  which includes Jones deciding to eat a blue sweet. This is consistent with, but does not require, the truth of a general thesis of causal determinism; as Frankfurt states, no deterministic relation is required to hold between the states of affairs at  $t_1$  including Jones’s twitch and any earlier states. If some such deterministic causal relation did obtain, then by intervening in all and only those scenarios in which Jones did not

---

<sup>114</sup> Frankfurt (in his 1971 article) agreed with Hobbes that an agent’s will is identical with at least one of their ‘first-order’ desires to do something. But, he distinguished these first-order desires from higher order desires – desire *about* one’s lower order desires. Higher order desires about which first-order desires motivate one’s actions were called, by Frankfurt, ‘volitions’. On his view, one would be ‘*acting of one’s own free will*’ only if one’s volitions match with one’s will (i.e. one’s will is desired at this higher level). One would have ‘*freedom of the will*’ only if, in addition, one is (or was) free to make some other first-order desire one’s will (i.e. only if one was free to *will* otherwise than one did). On Frankfurt’s view, an agent’s responsibility for some action entails that they acted freely, and that they acted of their own free will, but does not entail that they had freedom of action, nor that they had freedom of the will.

<sup>115</sup> Frankfurt, 1969, p. 6

<sup>116</sup> The twitch is supposed to be involuntary and not itself an exercise (or even the beginning of an exercise) of the agent’s will. The significance of this alternative *possibility*, for PAP, is a matter of controversy; I discuss this later in the chapter (§9).

<sup>117</sup> Frankfurt, 1969, p. 6 (fn.)

<sup>118</sup> Ibid. p. 6 (fn.)

twitch at  $t_1$ , Dr. Black would ensure that Jones will not (even begin to) decide at  $t_2$  to do anything other than eat a blue sweet.

A leeway compatibilist who accepts that agents are able to perform acts of will, and that their doing so is relevant to their freedom and responsibility, must accept that agents may retain this ability even if their decision is causally determined by prior states of affairs and the actual laws of nature. The assumption of a deterministic relation between twitch and decision is consistent, on this view, with the following claims: (a) Jones is responsible for what he has done, and (b) Jones could have done otherwise, in some relevant sense. In the actual scenario, Dr. Black was present, Jones did twitch at  $t_1$ , and Jones decided at  $t_2$  to eat a blue sweet. Assuming the relevant deterministic causal relation, it was nomologically impossible, given the precise conditions at  $t_1$ , for Jones to have decided otherwise at  $t_2$ . And, in no other nomologically possible scenario in which Dr. Black is present would Jones have decided otherwise at  $t_2$ . Hence, leeway compatibilists should agree that Jones could not have done otherwise, in any relevant sense.<sup>119</sup> Yet, they have reason to consider Jones to be responsible for what he has done. In a similar scenario in which Dr. Black is absent, Jones would, on their view, be responsible for what he has done; and, in the actual scenario, Dr. Black played no actual causal role in bringing about Jones's decision and action. I take it, then, that this account of Dr. Black's powers gives Frankfurt a counterexample against initially plausible leeway compatibilist interpretations of PAP.

The assumption of a deterministic relation between states including Jones's twitch and his subsequent decision is, however, not obviously dialectically innocent, insofar as Frankfurt-style cases challenge leeway *incompatibilism*. To claim that Jones would be responsible for what he has done in such a scenario would be to assume that agents may be responsible for what they have done, even if their decision and action were causally determined, given past states in conjunction with the laws of nature. While this is acceptable for a leeway compatibilist, this is precisely what leeway incompatibilists oppose. To assume, as part of an argument against leeway incompatibilism, that responsibility is compatible with determinism, would be question-begging.<sup>120</sup>

### 3. The Dilemma Defence of Leeway Incompatibilism

A dilemma threatens Frankfurt's strategy, insofar as his example is supposed to challenge leeway incompatibilism.<sup>121</sup> If Frankfurt is not dialectically entitled to claim both that a deterministic relation obtains between prior states and Jones's decision, *and* that Jones is responsible for what he has done in

---

<sup>119</sup> I have argued above that even if we grant that Jones retained, in this scenario, a masked ability to have tried (and/or decided, and/or chosen) to act otherwise, this is not the sort of ability which is proposed (at least by plausible leeway compatibilist theories) to be necessary for Jones to be responsible for what he has done.

<sup>120</sup> Frankfurt is aware of this – see his 1969, pp. 6 – 7 (fn. 3)

<sup>121</sup> For examples of this style of argument, see Widerker, 1995, p. 247, Ginet, 1996, pp. 406 – 408, and Kane, 1996, p. 142

this case, then he must suppose that the relation between prior states and Jones's decision is indeterministic. If so, and if Dr. Black maintained his plan to intervene when and only when Jones does not twitch at  $t_1$ , his presence seems to justify, at most, the claim that Jones will *probably* not decide otherwise at  $t_2$ ; this model of intervention would allow Jones to retain, in Black's presence, whatever ability to do otherwise, having twitched, he had in Black's absence. Either way, Frankfurt has not provided leeway incompatibilists with a counterexample to their PAP.

#### 4. Responding via the 'Deterministic Horn'

Anyone who asserts the incompatibility of determinism and responsibility must deny that, in any Frankfurt-style case which assumes a deterministic relation between some prior (involuntary) event and the agent's decision, the agent is responsible for what they have done. If it were claimed that the agent in such a Frankfurt-style case is responsible for what they have done, then (in the absence of independent argument in its favour) the compatibility of determinism and responsibility would be illegitimately assumed, and the proposed counterexample would be question-begging. But, it does not follow, from the claim that the agent in such a case is *not* responsible for what they have done, that the agent's non-responsibility is due to the incompatibility of determinism and the ability to do otherwise and, therefore, their non-satisfaction of this necessary condition for responsibility. A *leeway* incompatibilist would make such explanatory claims. But, this explanation of the incompatibility of determinism and responsibility is not mandatory for incompatibilists; *source* incompatibilists, for example, consider determinism to be incompatible with responsibility for reasons other than the proposed incompatibility of determinism and the ability to do otherwise.

A deterministic-type Frankfurt-style case might, then, count against *leeway* incompatibilism if it could show that the agent's inability to do otherwise does not explain their *non*-responsibility for what they have done. This would not require the assumption of the compatibility of determinism and responsibility; it would leave this question open, and merely constrain the possible explanations of incompatibility.<sup>122</sup> This style of indirect argument has been proposed by Fischer, who claims that it succeeds in undermining leeway incompatibilism, via deterministic-type Frankfurt-style cases, while avoiding the allegation of question-begging.<sup>123</sup>

Fischer has argued that Dr. Black's presence in such cases (a) explains, in itself, and apart from the assumption of causal determinism, why Jones could not have decided otherwise at  $t_2$ , and (b) is irrelevant to explaining Jones's (non)-responsibility for choosing and acting as he did – Dr. Black does not

---

<sup>122</sup> It might *then* be argued that determinism – aside from threatening the ability to do otherwise – does not directly threaten responsibility.

<sup>123</sup> See Fischer 1999, p. 113, his 2002 (in Kane (ed.) 2002) pp. 291 – 292, and his 2010, pp. 323 – 334.

intervene, and plays no role in actually determining Jones's decision.<sup>124</sup> So, he has drawn the following initial conclusion: if responsibility is incompatible with causal determinism, this is not because causal determinism would ensure that no agent could have done otherwise.

It seems, however, that this conclusion is unwarranted. Fischer's argument is made in response to the suggestion that the assumption of a deterministic relation between prior-sign and decision is *necessary* for Dr. Black to ensure that Jones could not have decided otherwise at  $t_2$ ; by taking the deterministic horn of the dilemma, we may suppose that Fischer has conceded this point.<sup>125</sup> So, we have reason to suppose that (a) is false; without the assumption of determinism, Dr. Black's presence would fail to explain why Jones could not have decided otherwise at  $t_2$ . And, as leeway incompatibilists claim that the assumption of determinism is sufficient, in itself, to ensure that Jones could not have decided otherwise at  $t_2$ , Dr. Black's presence seems to be explanatorily redundant: it seems that Jones could not have decided otherwise at  $t_2$  because he was actually determined to decide at  $t_2$  as he did, whether or not Dr. Black was also present and ready to intervene.<sup>126</sup>

If this is correct, then Fischer is not entitled to assert (b); whereas Dr. Black (who did not actually intervene) may be irrelevant to explaining why Jones acted as he did (and, therefore, his (non)-responsibility for having done so), the deterministic relation actually obtained and determined Jones's behaviour. This would, then, be a case in which the *same* factor (i.e. the prior determination of his decision) explains both (i) Jones's deciding as he did, and (ii) the fact that he could not have decided otherwise; Frankfurt's case was supposed to show that a *different* factor from that which explains (i) may explain (ii).<sup>127</sup> By failing to distinguish these factors, Fischer fails to show that the obtaining of the deterministic relation, which explains why Jones could not have done otherwise, does not thereby explain why Jones is non-responsible for what he has done.

Fischer has argued that, even if the deterministic relation is sufficient, in itself, to explain why Jones could not have decided otherwise at  $t_2$ , this does not entail that nothing else could be a partial or complete explanation of this fact.<sup>128</sup> He cites, as an example, the possibility of multiple complete explanations of the same phenomena, given at different levels; for example, psychological phenomena may be subject to complete explanation at the sub-personal level, given in mechanistic terms, and complete explanation at the personal level, given in terms of mental states.

---

<sup>124</sup> Fischer, 2010, pp. 324 – 325

<sup>125</sup> If he has a compelling argument showing that Black may have this power, under the assumption of an indeterministic relation between prior states and decision, then he has already vindicated Frankfurt's argument, and need not attempt to construct a successful case under the assumption of a deterministic relation (see Garnett, 2013, p. 4).

<sup>126</sup> See Goetz, 2005, pp. 87 – 89

<sup>127</sup> Frankfurt, 1969, p. 1

<sup>128</sup> Fischer, 2010, pp. 332 – 333

Presumably, Fischer's thought is that we may grant that the deterministic relation between prior sign and events correlated with the decision is a complete mechanistic sub-personal level explanation of the fact that Jones could not have decided otherwise at  $t_2$ , but that Dr. Black's presence is a complete non-mechanistic personal level explanation of this fact. But, if we describe the case at the personal level, in terms of the agents' psychological states, then we must include determinism – it must be the case that Black has the power to predict and pre-empt Jones's decision, and as we have taken the deterministic horn, we are supposing that Black can do so only if a deterministic relation obtains between the relevant psychological states of Jones.<sup>129</sup> There seems to be no additional reason to suppose that, at this level, Black's presence plays some role in explaining why Jones could not have decided otherwise; moving to this level of explanation does not resolve the original problem. I take it, then, that Fischer's argument is unsuccessful, and we have no reason to suppose that Frankfurt-style cases can succeed on the deterministic horn of the dilemma.

### 5. Responding via the 'Indeterministic Horn'

If this is right, then it must be assumed that Jones's decision is undetermined, given prior (involuntary) events. If so, and if Dr. Black were to maintain his policy of non-intervention on observing Jones's twitch at  $t_1$ , then given conditions at  $t_1$  and the laws of nature, it would be *possible* for Jones to go on to decide at  $t_2$  to act otherwise.<sup>130</sup> Jones *could*, then, in this sense, decide at  $t_2$  (or, presumably, before  $t_2$ ) to act otherwise.

The mere nomological possibility, given conditions at  $t_1$ , that Jones will go on to decide at  $t_2$  to act otherwise is, on any leeway incompatibilist account of freedom (of which I am aware), considered to be necessary but insufficient for Jones to be able to do otherwise in the PAP-relevant sense. Different leeway incompatibilist accounts of the relevant ability propose different additional necessary conditions.

### 6. Leeway Incompatibilist Theories of Freedom

Many leeway incompatibilists consider the power to act (including the power to perform mental actions, like deciding), and therefore the power to act freely, to be a causal power; on these views, the agent has the relevant sort of control over which open alternative they take (e.g. whether they decide to do A, or to

---

<sup>129</sup> Garnett, 2013, p. 6

<sup>130</sup> If there were cases in which Jones does not twitch at  $t_1$ , and would have gone on to decide in accordance with the plan at  $t_2$ , Dr. Black – on his current policy – would intervene unnecessarily. But, this is not a problem for Frankfurt; he requires only that Black intervenes in *all* cases in which Jones would otherwise have decided otherwise; he does not require that Black intervenes and *all of, and only*, those cases (see Hunt, 2000, p. 207).

do B, etc.) only if it is possible (given initial conditions) that they (or events involving them) *cause* the event of their taking alternative A, or the event of their taking alternative B, (etc.).<sup>131</sup>

Two views of this sort should be distinguished. On an *event-causal* view, a free decision is considered to be an event suitably caused by certain sorts of agent-involving events – characteristically, these are events in the agent’s psychology, such as their having certain beliefs and desires.<sup>132</sup> On leeway incompatibilist event-causal views, free decisions must be non-deterministically caused; some stage of the causal process must be indeterministic. The most popular views of this sort, and those on which I shall focus, are ‘centred accounts’, according to which it is required that the immediate causal antecedents of the decision (i.e. the relevant psychological events) indeterministically cause the decision.<sup>133</sup>

On an *agent-causal* view, a free decision is considered to be an event caused by an *agent*, and causation by an agent does not consist in causation by events.<sup>134</sup> An agent is said to be a persisting substance, whereas events (including events which involve the agent) are merely momentary occurrences. Agents, as persisting substances, are not the kind of thing which could be effects of other events, but are, according to advocates of these views, things which possess causal powers; insofar as agents exercise these causal powers, their doing so must be undetermined by prior events. On these views, agents have the power to determine what they do, without prior determination; they may be the uncaused cause of what they do.

On an event-causal view, it is necessary, for Jones to be able to do otherwise in the relevant sense, that (i) Jones’s decision is not determined by prior events and actual laws, and (ii) certain events within or involving Jones, given the precise conditions, might cause Jones to decide to act otherwise. And, on an agent-causal view, it is necessary, for Jones to be able to do otherwise in the relevant sense, that (i) Jones’s decision is not determined by prior events and actual laws, and (ii) Jones has the power to agent-cause his decision to act otherwise. It was granted, when the Frankfurt-style case was set up, that (in the intervener’s absence) Jones could have done otherwise, in whatever sense one considers to be necessary for responsibility; given this, Dr. Black’s presence, and capacity to intervene as described above (i.e. he

---

<sup>131</sup> Some consider the relevant power to be a non-causal power (e.g. McCann, 1998, and Pink, 2004, pp. 116 – 123). These views face the challenge of accounting for agent’s control over their actions, and of their acting for reasons, without appealing to causation of the action by (states of) the agent. Many such views invoke the notion of intention, such that one has control, and acts for a reason, when one’s behaviour occurs concurrent with, or has as its content, a certain intention of one’s, which plays no role in bringing about (causing) one’s behaviour. There are various difficulties with such views (see Clarke & Capes, 2013, §1) so I set them aside here, but I do not reject them; for all that I go on to say, Frankfurt-style cases might be ineffective against such non-causal leeway incompatibilist views.

<sup>132</sup> Clarke, 2011, p. 337 (although Clarke does not advocate such a view).

<sup>133</sup> See, for example, Kane, 1996, pp. 100 – 120. (On other ‘deliberative’ event-causal views, it must be undetermined which of the agent’s beliefs and desires come to mind in deliberation; the ensuing action-producing causal process is, then, allowed to be deterministic - see, for example, Mele, 1995 (c. 12), or Ekstrom, 2000, p. 106. If the ‘coming to mind’ of certain beliefs and desires is beyond the agent’s control, and this determines their subsequent decision, it is difficult to see how such an agent could be acting consistent with a plausible incompatibilist freedom (see Clarke & Capes, 2013, §2.1). I set these views aside here.

<sup>134</sup> See, for example, Clarke, 2005, p. 409

will intervene when and only when Jones does not twitch at  $t_1$ ) is consistent with Jones being able to do otherwise, in the relevant sense, according to both of these leeway incompatibilist views.

### 7. Prior-sign Intervention and Event-causal Views

For example, the following scenario is consistent with the case as described above. Suppose that a twitch of the sort which occurred in Jones at  $t_1$  invariably occurs in him when he holds a particular set of beliefs and pro-attitudes; Jones twitches when he recognises what he considers to be a strong reason to eat a blue sweet, and also recognises what he considers to be a weaker reason to eat a red sweet. Call the event of his recognising what he considers to be a strong reason to eat a blue sweet, in these circumstances, event 'C'. The occurrence of event C satisfies the stimulus conditions of a probabilistic disposition of Jones's, the manifestation of which is the event of Jones's decision, at a specified later time, to eat a blue sweet (call this event 'E'). Whenever event C occurs, there is a probability  $n$  (greater than 0 and less than 1) that this event will cause event E to occur at this later time, and there is a probability of  $1-n$  that event C will not cause event E to occur.<sup>135</sup> Suppose that Jones did decide at the specified later time ( $t_2$ ) to eat a blue sweet. In this case, the occurrence of event C non-deterministically caused event E to occur; up until it did, there was a chance that Jones would not decide to eat a blue sweet, or that he would decide otherwise.

Suppose instead that this is a case in which event C does not cause event E to occur, and Jones decided otherwise (e.g. to eat a red sweet) at  $t_2$ . In this case, the occurrence of some *other* prior event non-deterministically caused him to make this decision.<sup>136</sup> For example, we may also suppose that Jones has a second probabilistic disposition. Call the event of Jones's recognising what he considers to be a reason to eat a red sweet, in these circumstances, event 'D'. The stimulus conditions of this second disposition are satisfied when event D occurs, and the manifestation of this disposition is the event of Jones's decision, at a specified later time, to eat a red sweet (call this event 'F'). Whenever event D occurs, there is a probability of  $n$  (greater than 0 and less than 1) that this event will cause event F to occur at this later time, and there is a probability of  $1-n$  that event D will not cause event F to occur. In this case, the occurrence of event D (i.e. Jones's recognition of what he considers to be a reason (albeit a weaker reason) to eat a red sweet) would have non-deterministically caused him to decide as he did.

So, if we suppose that Dr. Black observed Jones's twitch at  $t_1$  and therefore refrained from intervening, and that Jones went on to decide and act in accordance with his (Black's) plan, Dr. Black would not – at least according to these event-causal leeway incompatibilists – have ensured that Jones could not have done otherwise, in a relevant sense. The probabilistic disposition of Jones to decide to act otherwise –

---

<sup>135</sup> Clarke, 2011, p. 337

<sup>136</sup> See Clarke & Capes, 2013, §2.2

given Dr. Black's presence and capacity to intervene – is retained; in fact, it was not manifested, but it *could* have been manifested, in this scenario.

Given the occurrence of the twitch at  $t_1$ , even if we suppose that it was extremely probable that event  $C$  would cause event  $E$  to occur (at  $t_2$ ), up until it actually did so there was a chance that event  $D$  could have caused event  $F$  to occur, or at least to begin. The occurrence of the twitch at  $t_1$  does not provide Dr. Black with evidence that he need not intervene. It seems that he is left with two options: (i) to intervene indiscriminately, in which case he would, in all cases, determine that Jones decides and acts as he (Black) wishes, thereby failing to provide a case in which a *different* factor from that which explains why Jones could not have done otherwise explains why Jones acted as he did (and this is required for a successful Frankfurt-style case), or (ii) to intervene after the occurrence of event  $F$ , or after it has begun. Insofar as we suppose that the beginning of event  $F$  would amount to the beginning of an alternative decision, then this kind of intervention would fail to ensure that Jones could not have done otherwise in any relevant sense.<sup>137</sup> Given this supposition, neither option promises to provide a successful Frankfurt-style case.

## 8. Blockage Cases

Some have sought to avoid such difficulties by re-formulating the Frankfurt-style case, such that the intervention is not triggered by a prior-sign.

### (a) Hunt's Case

For example, Hunt has noted an anticipation of Frankfurt's argument in the writing of John Locke; he (Hunt) presents a modified version of a case described by Locke, and claims that it avoids the above problems.<sup>138</sup>

In Locke's case a free agent is placed inside a room whilst sleeping; the room is then locked, and some time later the agent awakes to find something of interest to him.<sup>139</sup> He then remains in the room willingly, examining the thing of interest, unaware that, had he tried to leave the room, he would have failed to do so. It seems that he remained in the room freely, and is responsible for having done so, even though he could not have left the room. Of course, the agent could have decided to leave the room, and made an overt attempt to do so; PAP is not threatened by this case.<sup>140</sup>

---

<sup>137</sup> As I go on to discuss, it is a matter of some significance whether or not the beginning of some such event counts as (at least the beginning of) an exercise of the agent's will.

<sup>138</sup> Hunt, 2000, p. 217

<sup>139</sup> *Essay Concerning Human Understanding*, Bk.II, C.XXI, §10

<sup>140</sup> This is not to say that Locke failed to show what he intended to show, via this case.

Hunt proposes that analogous intervention can ensure that Jones could not even have decided to do otherwise; ‘blockages’ (analogous with Locke’s door) may be introduced within the agent’s decision-making system, such that they cannot decide otherwise, but nonetheless can decide ‘on their own’ to perform some action. We are to suppose that at some time prior to his deliberation, all of Jones’s ‘neural pathways’, apart from one, are blocked (by Dr. Black); neural firings which follow this open pathway somehow correlate with Jones deciding, at  $t_2$ , to eat a blue sweet. The firings in Jones’s brain follow this pathway, and he makes this decision, but Hunt claims that the blockages may be causally irrelevant; it may be the case that, had there been no blockages, the firings in Jones’s brain would have followed precisely this pathway.<sup>141</sup> In this case, he claims that we have a successful Frankfurt-style case, without reliance on a prior-sign.

But, it is unclear how the presence of such ‘blockages’ is supposed to ensure that Jones could not have done otherwise, in any relevant sense. If, given conditions at  $t_1$  (at which point, the blockages are in place) and the laws of nature, the presence of the blockages ensures that only one event is physically possible at any time between  $t_1$  and  $t_2$ , then their presence determines Jones’s decision at  $t_2$  and this case must face the deterministic horn of the original dilemma. The presence of the blockages must, then, be consistent with the possibility that other events could occur prior to, and at,  $t_2$ . Yet, if this is conceded, we have no reason to suppose that neural firings might not diverge from the specified path, only to meet with the blockage and thereby be caused to return to the specified path.<sup>142</sup> If so, there would remain a possibility that event  $D$  (Jones’s recognising a reason to eat a red sweet) would cause some divergence towards the causation of event  $F$  (his deciding to eat a red sweet), before the causal process is cut short and a different deterministic process is induced. Without the assumption of blockage-induced determinism, this case is exactly analogous with the case described above, in which Dr. Black intervenes only *after* some initial divergence. On the supposition that such divergence would amount to an exercise of Jones’s will, I have argued that this kind of intervention does not succeed.

#### (b) Mele & Robb’s Case

Mele and Robb have attempted to ‘rescue’ Frankfurt-style cases from the original dilemma via the following modified case.<sup>143</sup> Prior to Jones’s deliberation, Dr. Black has induced a deterministic causal process ‘ $P$ ’ in Jones’s brain, such that he will decide at  $t_2$  to eat a blue sweet, unless (a) he decides on his own at, or prior to,  $t_2$  to eat a blue sweet, or (b) is incapable of making a decision at  $t_2$  (e.g. he is unconscious or dead). In fact, Jones does decide at  $t_2$ , via his own indeterministic decision-making process ‘ $U$ ’, to eat a blue sweet, and the operation of process  $P$  does not affect the operation of process  $U$ .

---

<sup>141</sup> Hunt, 2000, p. 218

<sup>142</sup> Fischer calls these ‘bumping’ events – see his, 1999, p. 119

<sup>143</sup> Mele & Robb, 1998, pp. 101 – 102

For this case to succeed, two things must be established. Firstly, why is it only process *U* which is causally effective in bringing about Jones's decision at  $t_2$ , and not process *P*? If process *P* were causally effective, then the case would face the deterministic horn of the original dilemma. And, secondly, if it is granted that Jones is able to bring about his decision at  $t_2$  via process *U* only, why is Jones unable to decide otherwise via process *U* at, or before,  $t_2$ ?

With regard to the first point, we are asked to suppose that it is necessary, for one to make a decision, that a 'decision-node' in one's brain is caused to 'light up'.<sup>144</sup> Normally, events in Jones's psychology non-deterministically cause these nodes to light up (this is 'process *U*'). Prior to  $t_1$ , Dr. Black has initiated a deterministic causal process (i.e. 'process *P*') which includes the relevant node (i.e. the node, the lighting of which is necessary for a decision to eat a blue sweet) lighting up at  $t_2$ , and which leaves Jones's normal decision-making process intact. Significantly, the relevant node is sensitive to the kind of event which is about to cause it to light up; if both an event in process *U* and an event in process *P* were about to cause it to light up, at  $t_2$ , the node would light up only because of the event in process *U*. Causal process *U* is said to trump causal process *P* when they converge on this node at  $t_2$ .

If this is granted, then might process *U* have this power when striking other nodes at  $t_2$ ? If so, Jones – via process *U* – could decide, at  $t_2$ , to act otherwise. Perhaps it could be argued that the lighting up of the relevant node would trump the simultaneous lighting up of any other node (regardless of the process which caused it to light up) in the subsequent causation of Jones's decision. This would, however, leave open the possibility that process *U* be about to cause an alternative node to fire, only for this process to be cut short by the activity of process *P*. In fact, Mele and Robb suggest that all other nodes are 'neutralised' when Black initiates process *P*.<sup>145</sup> Presumably, this means that if process *U* were to arrive at any other nodes at or before  $t_2$ , it would not cause them to light up. But, similarly, this leaves open the possibility that process *U* meets with a blockage at or before  $t_2$  and is rendered ineffective. It seems, then, that this case is very similar to Hunt's; it depends on prior blockage of processes necessary for alternative decisions, but permits alternatives of a certain sort. Insofar as we suppose that such alternatives would amount to an exercise of Jones's will, we have no reason to suppose that Jones could not have done otherwise, in some relevant sense, in this case.

## 9. Flickers of Freedom

The difficulty with these modified Frankfurt-style cases is that, in order to avoid the deterministic horn of the original dilemma, they must allow that, given the presence of the intervener or mechanism, alternative events of some sort may occur. And, for all that has been specified in these cases, we have no reason to

---

<sup>144</sup> Ibid. pp. 103 – 104

<sup>145</sup> See *ibid.* p. 105 (fn. 16)

suppose that these alternative events might not amount to (at least the beginning of) an alternative act of will. The intervention would, therefore, fail to ensure that the agent could not have done otherwise in some relevant sense.

One apparent advantage of intervention triggered by a prior-sign, over blockage-style intervention, is that it can be made clearer that the possibility of the occurrence of the triggering event (e.g. the event of Jones not twitching at  $t_1$ ) does not suffice for Jones to be able even to begin an alternative act of will.<sup>146</sup> The triggering aspect of prior-sign cases may be consistent with the falsity of general causal determinism, as they may allow that given the precise conditions up until  $t_1$ , and the laws of nature, it is possible for Jones to twitch and for him not to twitch at  $t_1$ . They may, therefore, allow alternative *possibilities*; a different event, involving Jones's body, might have occurred at  $t_1$ , thereby triggering intervention. But, mere possibilities of this sort do not (or, at least not obviously) suffice for Jones to be able to do otherwise, in any relevant sense.

For example, it may be specified that the occurrence of the event of Jones not twitching at  $t_1$  (a) would not itself amount to the beginning of a decision to act otherwise and (b) would have been beyond Jones's control (i.e. the twitch, or non-twitch, would have been involuntary).<sup>147</sup> In such a case, Jones could, in some sense, have done otherwise at  $t_1$ : he might not have twitched (and then been caused, via Dr. Black's intervention, to act in accordance with the plan). But, if conditions (a) and (b) are satisfied by the triggering event, it seems that Jones satisfies no plausible version of PAP; he 'could' have done otherwise only in a very weak sense – of mere *possibility* – rather than in the stronger sense – of *ability* – which seems to be required.

Fischer has called alternative possibilities of this sort mere 'flickers of freedom', and has argued that the fact that such 'flickers' remain in a Frankfurt-style case is insufficient to show that this case does not count against PAP.<sup>148</sup> According to Fischer, alternative possibilities are relevant only insofar as they are "...sufficiently robust to ground our attributions of...responsibility".<sup>149</sup> And, he has claimed that it is "highly implausible that it is in virtue of the existence of such an alternative possibility [e.g. of Jones not twitching at  $t_1$ ] that Jones is...responsible for what he does".<sup>150</sup>

---

<sup>146</sup> Perhaps the blockage cases could be modified, such that the alternative processes earlier in the causal chain are also blocked. But, this would lead to difficulties regarding the prior determination of the action, analogous with those faced by prior-sign cases; blockage would, then, lose its initial appeal.

<sup>147</sup> If involuntariness is not specified (i.e. if it is allowed that the occurrence of the twitch at  $t_1$  is 'up to' the agent), then even if a deterministic relation is assumed to hold between events at  $t_1$  and the ensuing decision at  $t_2$ , one might claim that the agent is responsible for what they have done, without threatening PAP; this may be a case of indirect (or derivative) responsibility.

<sup>148</sup> Fischer, 1994, pp. 140 – 147 (see also his 2011, pp. 260 – 261).

<sup>149</sup> Fischer, 1999, p. 110

<sup>150</sup> Fischer, 1994, p. 140

Fischer has argued along the following lines.<sup>151</sup> Suppose that, given the precise conditions prior to  $t_1$ , and the actual laws of nature, it is impossible for Jones not to twitch at  $t_1$ , and that his twitch is deterministically related to his decision at  $t_2$ . Any leeway incompatibilist will deny that Jones is responsible for deciding as he does at  $t_2$ . But, now suppose that it was possible, given prior conditions and the laws, for Jones to have twitched at  $t_1$  and then to have been determined (by an intervention) to make a decision at  $t_2$  of the same type (e.g. to eat a blue sweet). Normally, the presumption in a Frankfurt-style case is that Jones *is* responsible for what he has done in the actual scenario, and alternatives are *removed* until only ‘flickers’ remain. Those Fischer calls ‘flicker theorists’ argue that it is in virtue of these flickers that Jones is responsible for what he has done in the actual scenario. But, in this case, the presumption is that Jones is *not* responsible for what he has done in the actual scenario, and exactly analogous ‘flickers’ are *added*. Flicker theorists assume the burden of explaining why, in virtue of these added flickers, Jones might be responsible for what he has done in the actual scenario. According to Fischer, they cannot do so without appealing to some kind of ‘alchemy’; he concludes that, if Jones is responsible, in the Frankfurt-style case, for what he has done in the actual scenario, then this is not in virtue of the presence of flickers.

There is, however, a sense in which such flickers do seem to ‘play a role’ in grounding attributions of responsibility. We may suppose that the event of Jones not twitching at  $t_1$  is (in the absence of an intervener) deterministically related to his deciding at  $t_2$  to act otherwise. So, in the alternative scenario Jones would have decided as he did only because of the intervention. There is, then, an alternative possible scenario in which Jones avoids authorship of, and responsibility for, what he has done; by being such that he would otherwise have decided otherwise at  $t_2$ , Jones (in some sense) made it the case that Dr. Black is the author of, and therefore responsible for, what Jones has done.<sup>152</sup> And, perhaps this possibility is sufficiently ‘robust’ to account for Jones’s responsibility for what he has done in the actual scenario.

Fischer has granted that these Frankfurt-style cases leave open an alternative possible scenario in which the agent avoids authorship of, and responsibility for, what they have done. But, he has argued that they do not thereby leave open ‘robust’ alternatives, because the agent cannot *voluntarily* avoid responsibility; it is possible for them to avoid responsibility only ‘accidentally’.<sup>153</sup> If the occurrence of the triggering event is not something over which the agent has voluntary control, then it does not amount to an act of the agent’s will, or even the beginning of such an act. Fischer has noted a parallel between the debate over the significance of ‘flickers’ and the relationship, which I mentioned earlier, between freedom and

---

<sup>151</sup> See Fischer, 1999, pp. 110 – 111

<sup>152</sup> See Wyma, 1997, p. 57, and McKenna, 1997, pp. 74 – 75. Otsuka (1998, pp. 691 -692) offers a slightly different response, which is relevant only to ascriptions of blameworthiness. Otsuka proposes a necessary condition for blameworthiness (i.e. PAB: ‘one is blameworthy for what one has done only if one could have behaved less badly’) and claims that either (i) the agent in a Frankfurt-style case retains the ability to behave less badly (e.g. they could have acted from more worthy motives) or (ii) if they lack this ability, then they are not blameworthy. Due to the comparatively narrow scope of this argument, and various difficulties involved in motivating and defending PAB, I will not consider this response any further.

<sup>153</sup> Fischer, 1999, p. 121

indeterminism: those committed to leeway incompatibilism must account for the “crucial difference between the ability to do otherwise, and the mere possibility of something different happening”, and it seems that mere possibility of the occurrence or non-occurrence of a twitch is insufficient to indicate the agent’s retention of an *ability* to do otherwise, even if it would determine their responsibility for their ensuing behaviour.<sup>154</sup> If these are the only possibilities at  $t_1$ , then it seems that no further necessary condition (for the agent’s retention of this ability) can be met.

Of course, the model of intervention on which it is triggered by a prior-sign leads to difficulties already discussed regarding the relationship between prior-sign and decision; if a deterministic relation is assumed, it becomes problematic to suppose that the agent acted freely and is responsible for what they have done, and if an indeterministic relation is assumed, it becomes problematic to suppose that the agent could not have done otherwise, in some relevant sense. These prior-sign cases do, however, seem to be preferable in at least one way; their structure seems to allow agents only mere ‘flickers’ of freedom, rather than the genuine freedom to do otherwise which I argued – on the supposition that initial divergences from the specified paths were *voluntary* – is permitted by the blockage cases considered. If this is right, a prior-sign case in which conditions (a) and (b) are satisfied by the triggering event, in which this ‘flicker’ is the *only* alternative possibility, and which (somehow) avoids the problematic assumption of prior determination of the agent’s act of will, would be a promising Frankfurt-style case.

#### 10. Stump’s Case

The following modified Frankfurt-style case, introduced by Stump, may fulfil these criteria.<sup>155</sup> Dr. Black is able to monitor Jones’s neural firings, and knows that any neural sequence which begins with neural firings  $a,b,c$ , is correlated with, and only with, a mental act of Jones deciding to eat a blue sweet.<sup>156</sup> He therefore knows that any neural sequence which begins with neural firings other than  $a,b,c$  is correlated with a mental act of Jones deciding to act otherwise. So, Dr. Black’s intervention would be triggered by his observation of the beginning of a neural sequence with firings other than  $a,b,c$ ; he would then induce, in Jones’s brain, a neural sequence beginning with firings  $a,b,c$ , determining that Jones perform his (Dr. Black’s) preferred action. In fact, Dr. Black observes neural firings  $a,b,c$ , and does not intervene; Jones decides to eat a blue sweet, and does so.

Dr. Black’s intervention would be triggered by firings other than  $a,b,c$ , and therefore by the commencement of a neural sequence correlated with an alternative act of will. For all that has been specified so far, one might object that the commencement of this sequence would be correlated with the

---

<sup>154</sup> Fischer, 2002 (reprinted in Watson (ed.), 2002) p. 196

<sup>155</sup> See Stump 1999, pp. 305 – 307

<sup>156</sup> It is assumed that there is some correlation between neural states and mental acts and states; this is compatible with various theories of the relation between mind and brain. See Stump, 1999, pp. 305 – 306.

beginning of an alternative act of will, and therefore that this case leaves the agent with ‘robust alternatives’ by Fischer’s standards: Jones retains the ability to begin to decide to act otherwise, and to be prevented from completing this decision. Stump supposes, however, that there is a one-many relation between a mental state or act and the neural firings correlated with it, such that there is no mental state or act (not even the early stages of one) without a *completed* correlated neural sequence.<sup>157</sup> If so, Stump would be entitled to claim that Dr. Black’s intervention, triggered by firings other than *a,b,c*, would precede (even the beginning of) an alternative act of Jones’s will; if there is no act (or even ‘movement’) of an agent’s will without the completion of the appropriate neural sequence, then the occurrence of a partially completed neural sequence cannot be within the control of, or amount to an exercise of, the agent’s will. Of course, this leaves open the possibility that some firings other than *a,b,c* might occur (i.e. those which would trigger the intervention). But, because such an occurrence must be involuntary, given the above supposition, this possibility is a mere ‘flicker’ by Fischer’s standards; it is insufficiently robust to satisfy a plausible version of PAP.

Even if Stump’s supposition is granted and it is conceded that the remaining alternatives are non-robust, a different objection may be pressed: if firings *a,b,c* precede the mental act and suffice to determine that it will be performed (i.e. that Jones will decide to eat a blue sweet), and if the occurrence of these firings is not within Jones’s control, then his act of will is determined by prior occurrences beyond his control, and therefore this case fails to avoid the deterministic horn of the original dilemma. According to Stump, however, nothing specified in this case is incompatible with the assumption that Jones acted indeterministically; it is not required, for successful intervention, that something “acted causally on Jones to produce the act in question”, because the intervention would be triggered by the early stages of a neural process correlated with the mental act in question, not by a prior event.<sup>158</sup> On an identity theory, for example, the completed sequence of neural events *is* the mental act.<sup>159</sup> Dr. Black may wait until the sequence has begun before judging whether to intervene; successful intervention is consistent with the indeterministic causation of this sequence, or with its uncaused occurrence.<sup>160</sup> On the assumption that the occurrence of the early stages of this sequence are beyond the agent’s voluntary control, this model of intervention would not leave open robust alternatives.

---

<sup>157</sup> Stump, 1999, pp. 306 – 307 (This might commit one to the view that a mental act is temporally extended. Identity theorists might be committed to this view – e.g. the mental act is identical with the (temporally extended) neural sequence, but (for some reason) the whole sequence of firings is required for the mental act to exist. On other theories, it might be claimed that the mental act exists only after the final neural firing in the sequence. See *ibid.* pp. 306 – 307 (fn. 20)).

<sup>158</sup> *Ibid.* p. 308

<sup>159</sup> More would need to be said about why the occurrence of the early stages of this neural sequence is not identical with the beginning of the mental act.

<sup>160</sup> *Ibid.* p. 310 – it might be more difficult to make a similar point on the assumption of some other view about the relation between mind and brain; for example, if mental events are distinct from, yet causally related to, physical events, perhaps the mental act might be determined to come into existence after, or simultaneous with, the completion of the sequence of physical events. As I wish to criticise Stump’s case on a different basis, I will not discuss this issue further.

## 11. Agent-Causation and the Power to Determine One's Own Will

Stump's case depends upon the assumption that the occurrence of any event in the neural sequence, prior to its completion, is beyond the agent's voluntary control. This assumption commits one to the view that it is not up to the agent whether a sequence beginning with firings  $a, b, c$  occurs, or a sequence beginning with, say, firings  $x, y, z$  occurs; these are events which occur in the agent's brain, independently of the agent's will. As the agent performs an act of will only when some such sequence of neural firings is completed, and lacks control over any part of this sequence prior to its completion, it seems clear that, on this picture, it is impossible for the agent to initiate and determine their own will; they must lack the power to bring about their own decisions (or, more broadly, acts of will (or 'volitions')).

This picture seems to be consistent with what is required for freedom on an event-causal leeway incompatibilist view. For example, the physical events correlated with psychological states of the agent (e.g. their holding certain beliefs and pro-attitudes) at some time ( $t_1$ ) might indeterministically cause the occurrence, in the agent's brain, of a neural sequence beginning with firings  $a, b, c$ , which is correlated with a decision to, say, eat a blue sweet. This leaves open the possibility that events correlated with the agent's holding certain other beliefs and pro-attitudes (which they also hold at  $t_1$ ) might indeterministically cause the occurrence, in their brain, of a neural sequence beginning with firings  $x, y, z$ , which is correlated with a decision to, say, eat a red sweet. In the absence of Dr. Black, the agent could, in the freedom-relevant sense (on this view), decide otherwise than to eat a blue sweet. But, given the presence of Dr. Black, and supposing the truth of Stump's assumptions, if a neural sequence beginning with firings  $x, y, z$  were to commence, it would not be completed. As the completion of the sequence is necessary for an alternative decision, then given conditions as they are in this scenario, no alternative decision could be made, or even begun.

Stump's case may, then, pose a problem for event-causal leeway incompatibilism. But, there is independent reason to suppose that such views do not provide an adequate account of the ability to do otherwise. Event-causal leeway incompatibilist theories of freedom face a 'problem of luck'. On these views, an agent is said to exercise freedom-relevant control over what they decide when their decision is indeterministically caused by the antecedent occurrence of certain agent-involving events. The agent's involvement in these antecedent events exhausts their role in bringing about their decision.<sup>161</sup> But, insofar as these events indeterministically cause the agent's decision, their occurrence must leave open the possibility that, given conditions precisely as they were up until the time of the decision, the agent might decide otherwise. So, for all the input that the agent has, it seems nonetheless to be a matter of chance whether, given their antecedent states, they will decide to eat a blue sweet, or decide otherwise. This, it is argued, is not sufficient for the agent to exercise freedom-relevant control; granted, the agent might have

---

<sup>161</sup> Pereboom, 2007, p. 102

decided otherwise, but it does not seem to be ‘up to them’ what they decide, in the sense that seems to be significant for freedom and responsibility.<sup>162</sup>

Suppose, for example, that the occurrence of these states were to indeterministically cause the agent’s decision to eat a blue sweet. In taking this decision, the agent would exercise a sort of agential control available under determinism (i.e. to be caused, via agent-involving events, to make this decision), with the addition of the open possibility that they exercise this same sort of control in the performance of a different decision.<sup>163</sup> Leeway incompatibilists must claim that the kind of agential control available under determinism is not sufficient for freedom of the relevant sort; so, it does not seem promising to suppose that the mere possibility of the occurrence of an alternative scenario in which this sort of control is exercised – a possibility which might obtain, but which is not within the agent’s power to bring about – could be sufficient for freedom of the relevant sort. The addition of indeterminism in the causal process seems to add nothing to the agent’s control, and seems insufficient for them to be able to do otherwise in a relevant sense.

A free agent, on an event-causal leeway incompatibilist view, is subject to luck in their performance of acts of will because their ‘freedom’ consists in *being* (indeterministically) *caused*, via events involving them (over the occurrence of which they lack control), to have volitions. Free agents, on such views, need not have the power to initiate their acts of will; hence, such views are consistent with the assumption made in Stump’s case. There is reason to suppose that *this* fact about event-causal views renders them vulnerable to the luck objection. Compatibilist theories of freedom tend to be event-causal theories; an agent’s freedom consists in their *being* (deterministically) *caused*, via agent-involving events, to have volitions. Incompatibilists must consider this kind of power to be insufficient for a kind of freedom worth wanting. But, to argue that this kind of freedom requires only indeterministic causation between agent-involving events and volitions seems nonetheless insufficient; this seems to concede too much to the compatibilist, namely that freedom consists in being caused (somehow) by agent-involving events to have volitions. If it is accepted that the agent need not have control over the occurrence of these agent-involving events, then no additional control – over and above that available on a compatibilist view – is added by the requirement of indeterministic causation.

Leeway incompatibilists may reject the claim that being caused (even in a way which is incompatible with the truth of determinism) by agent-involving events to have volitions is sufficient for the kind of freedom worth wanting (i.e. that which is relevant to agents’ responsibility). Some further power seems to be

---

<sup>162</sup> My comments in fn. 67 do not commit me to an analogous response here. Luck in the production of volitions seems to undermine control in a stronger sense than luck in the production of outcomes; the latter is compatible with, say, agent-causation of volitions.

<sup>163</sup> See Clarke, 2011, pp. 339 – 342

necessary, namely a power of agents to determine for themselves, without being caused to determine, what they will to do.

One way in which leeway incompatibilists have attempted to account for this sort of power is via an agent-causal theory; on these views, the relevant power is considered to be a power of the agent, as a persisting substance, to be the causal origin – the uncaused cause – of their volitions. These views seem to avoid the problem of luck.<sup>164</sup> Suppose that Jones has decided, at  $t_2$ , to eat a blue sweet. An agent-causal view requires that Jones's decision was undetermined; it must be the case that, given facts about the past and the laws of nature, Jones might have decided differently. It is consistent, then, with agent-causal views that there is a possible world, identical with the actual world up until  $t_2$ , in which Jones decides then to eat a red sweet. On an event-causal view, the difference between these worlds seems to be *only* a matter of luck. The agent's control is said to consist in the causation of volitions by agent-involving events, and there is no difference in the agent-involving events which occur in each world, while there is a difference in the volitions which occur in each world. Hence, for all of the agent's input, an alternative decision might or might not occur; the agent lacks the power to settle which probabilistic disposition will be manifested on this particular occasion. On an agent-causal view, however, the agent's control is said to consist in their power to cause their volitions, without being caused to do so; we may suppose that this power may be exercised in various ways, whatever agent-involving events precede its exercise. So, on this view, the difference between these worlds need not be only a matter of luck; in the alternative world, Jones may have exercised his power to cause a different volition to that which he caused in the actual world – it was within Jones's control to bring about each of these scenarios.<sup>165</sup>

If a leeway incompatibilist does deem some such power (i.e. to bring about one's own volitions) to be necessary for freedom of the sort required for responsibility, then their view will be inconsistent with the assumptions upon which Stump's case depends. Recall that it is granted, in the set-up of the Frankfurt-style case, that the agent is 'able to do otherwise' in the sense one deems to be necessary for responsibility. If this kind of power is deemed to be necessary, Stump's intervener cannot effectively intervene on the observation of the 'early stages' of a sequence of events correlated with an alternative decision, because the assumption that the occurrence of these events is beyond the control of the agent is unwarranted. If the agent exercises their power to initiate their act of will, no event which precedes their doing so counts as the 'early stages' of their act of will; if an act of will is correlated with a temporally extended sequence of events, then the occurrence or non-occurrence of even the earliest of these events will be, on this view, up to the agent. Hence, Stump's intervener, by waiting until some such sequence has commenced, will leave the agent with the ability to initiate an alternative act of will; this would be a robust alternative, by Fischer's standards, and the agent would, therefore, satisfy a plausible version of PAP.

---

<sup>164</sup> This is disputed by Mele, 2006 (see p. 70), and Haji, 2009 (see pp. 197 – 201). Clarke (2005) defends agent-causation against the luck objection.

<sup>165</sup> Clarke & Capes, 2013, §3.2

## 12. Conclusion

Stump's case promised to avoid the deterministic horn of the original dilemma, and to permit the agent only mere flickers of freedom. Contrary to other prior-sign cases, which either presuppose a deterministic relation between sign and decision, or leave open robust alternatives, intervention in Stump's case would be triggered by an event in the early, and involuntary, sequence of physical events correlated with the mental act of the decision. But, an incompatibilist of the sort under consideration need not concede that any such event could occur. The intervener's options are, then, constrained; they must find some other trigger for intervention. Their task is made even more difficult because, on this sort of incompatibilist view, even the smallest flicker could be relevant to PAP, as it would leave the agent able to do otherwise, in a relevant sense. So, for example, insofar as some alternatives are not 'blocked', even the smallest divergence from the path might constitute the beginning of an exercise of their will to act otherwise. And, if *all* alternatives are 'blocked', the case would face the deterministic horn of the original dilemma; I have argued that such cases do not succeed. It seems, then, that no Frankfurt-style case can succeed against leeway incompatibilism, insofar as leeway incompatibilists consider the ability bring about, without prior determination, one's own acts of will, to be necessary for freedom and responsibility.

If this conclusion is correct, it need not count against the effectiveness of Frankfurt-style cases against leeway incompatibilism. There might be independent reasons to reject the idea that it is necessary, for freedom of the relevant sort, that an agent possesses a power to bring about their own volitions, without determination by prior events, motives or other potential causes. This idea has much historical precedent in motivating commitment to incompatibilism.<sup>166</sup> But, it has also faced severe criticism.

According to Hobbes, such a power is unintelligible; in order for an agent to have such a power, their acts of will must be caused by prior acts of will.<sup>167</sup> For example, the agent must be able to decide whether they will make one decision or another; a vicious regress of prior decisions threatens. And, in any case, it seems that decisions are not the kind of thing which can be taken on the basis of prior decisions; if one decided to decide to perform some action, it would already be the case that one has decided to perform this action – after the initial decision, there would be no decision left to make.<sup>168</sup> Therefore, insofar as deliberate actions must be the effects of prior acts of will, the idea that one could 'will what one wills' seems to be incoherent. According to Hobbes, what one wills must be beyond one's voluntary control; he

---

<sup>166</sup> See Harris, 2005, p.7

<sup>167</sup> Hobbes, 'The Questions Concerning Liberty, Necessity and Chance', §1(c), (Chappell (ed.), 1999, p. 72).

<sup>168</sup> See Pink, 2004, p. 61 (and Clarke, 2009, p. 347) – One might decide to later *make a decision* about some issue, without deciding which option to take. But, one cannot decide to decide to do A, rather than B, without having decided to do A.

claimed that “I acknowledge this liberty, that I can do if I will, but to say that I can will if I will, I take to be an absurd speech”.<sup>169</sup>

Hobbes’s criticisms are not compelling. They depend on the assumption that, for something to be within one’s voluntary control, it must be the effect of a prior act of will. His opponents need not concede that this is a necessary condition for something like a decision to be within our control (‘up to us’) to perform.<sup>170</sup> Acts of will might be within our control in a different manner from that in which voluntary actions are within our control; they might, for example, be exercises of a distinctive capacity of ours.<sup>171</sup>

Other arguments purport to show that an agent with this sort of power would nonetheless lack freedom, given indeterminism.<sup>172</sup> If successful, these arguments would not show that this kind of power is not necessary for freedom; they would not, therefore, support Frankfurt’s argument. Even if it were established that this kind of power is physically impossible, we need not conclude that it is unnecessary for freedom or responsibility; this conclusion is consistent with the truth of hard (leeway) incompatibilism. If it could be shown that there is some conceptual flaw in the account of this power, such that its possession by human agents is impossible in some stronger sense, this might count against its necessity for freedom and responsibility.

The significant point, however, for the purposes of discussion here, is that Frankfurt-style cases have no bearing on this point.<sup>173</sup> Insofar as Frankfurt’s aim, in proposing this form of counterexample, was to provide an independently effective argument against leeway theory of *any* sort considered plausible by parties to the debate about free will and responsibility, then his strategy must fail. For Frankfurt-style cases to count against *all* plausible leeway incompatibilist views, it must be established independently that the ability to bring about, without prior determination, one’s acts of will, is not necessary for freedom. And, if this can be established, then – given much of the historical debate on this topic – a strong source of motivation for incompatibilism would be undermined already, independently of Frankfurt-style cases.

---

<sup>169</sup> Hobbes, ‘*Of Liberty and Necessity*’, §3, (Chappell (ed.), 1999, p. 16).

<sup>170</sup> Pink, 2004, p. 71

<sup>171</sup> See, for example, Pink, 2004, pp. 94 – 103

<sup>172</sup> See O’Connor, 2011, pp. 317 – 327

<sup>173</sup> Parts of the argument in this section are due to Elzein, 2009, pp. 109 – 119

## **Bibliography**

Austin, J.L. (1961) *Philosophical Papers* (Ed. Urmson, J.O. & Warnock, G.J.), Oxford University Press.

Anscombe, G.E.M. (1971) *Causality and Determination*, Cambridge University Press (reprinted (pp. 88 – 104) in Sosa, E & Tooley, M (eds.) (1993) *Causation – Oxford Readings in Philosophy*, Oxford University Press).

Ayer, A.J. (1954) 'Freedom and Necessity' (pp. 271 – 284) in *Philosophical Essays*, Macmillan, London (reprinted in Watson, G. (ed.) (1982) *Free Will* (First Edition), OUP, pp. 15 – 23).

Berofsky, B. (2011) 'Compatibilism Without Frankfurt: Dispositional Analyses of Free Will', (pp. 153 – 174) in Kane, R. (ed.) (2011) *The Oxford Handbook of Free Will* (Second Edition), Oxford University Press

Clarke, R. & Capes, J. (2013) 'Incompatibilist (Nondeterministic) Theories of Free Will', *The Stanford Encyclopedia of Philosophy* (Summer 2013 Edition), Zalta, E.N. (ed.), URL = <  
<http://plato.stanford.edu/entries/incompatibilism-theories/>>

Dennett, D.C. (1984) *Elbow Room – The Varieties of Free Will Worth Wanting*, Clarendon Press, Oxford.

Campbell, J.K. (1997) 'A Compatibilist Theory of Alternative Possibilities', *Philosophical Studies*, Vol. 88, No. 3 (Dec. 1997), pp. 319 – 330

Campbell, J.K. (2005) 'Compatibilist Alternatives', *Canadian Journal of Philosophy*, Vol. 35, No. 3, pp. 387 – 406

Chappell, V. (ed.) (1999) *Hobbes and Bramhall on Liberty and Necessity – Cambridge Texts in the History of Philosophy*, Cambridge University Press

Chisholm, R.M. (1964) 'Human Freedom and the Self', University of Kansas Lindley Lecture 1964 (reprinted in Watson, G. (ed.) (2003) *Free Will – Oxford Readings in Philosophy* (Second Edition), Oxford University Press, pp. 26 – 37).

- Choi, S. & Fara, M. (2013) 'Dispositions', The Stanford Encyclopedia of Philosophy (Summer 2013 Edition), Zalta, E.N. (ed.), URL = < <http://plato.stanford.edu/entries/dispositions/>>
- Clarke, R. (2005) 'Agent Causation and the Problem of Luck', *Pacific Philosophical Quarterly*, 86, pp. 408 – 421
- Clarke, R. (2009) 'Dispositions, Abilities to Act, and Free Will: The New Dispositionalism', *Mind*, Vol. 118, 470 (April 2009), pp. 323 – 351
- Clarke, R. (2011) 'Alternatives for Libertarians', (pp. 329 – 348) in Kane, R.H. (ed.) (2011) *The Oxford Handbook of Free Will* (Second Edition), Oxford University Press
- Ekstrom, L.W. (2000) *Free Will: A Philosophical Study*, Boulder: Westview Press.
- Elzein, N. (2009) *Freedom of the Will: A Possible Alternative*, Doctoral Thesis, University College London.
- Fara, M. (2005) 'Dispositions and Habituals', *Nous*, Vol. 39:1, pp. 43 – 82
- Fara, M. (2008) 'Masked Abilities and Compatibilism', *Mind*, Vol. 117, No. 468 (Oct. 2008), pp. 843 – 865
- Fischer, J.M. (1982) 'Responsibility and Control', *The Journal of Philosophy*, Vol. 79, No. 1 (Jan. 1982), pp. 24 – 40
- Fischer, J.M. (1994) *The Metaphysics of Free Will*, Blackwell.
- Fischer, J.M. (1999) 'Recent Work on Moral Responsibility', *Ethics*, Vol. 110, Oct. 1999, pp. 93 – 139
- Fischer, J.M. (2002) 'Frankfurt-type Examples and Semi-Compatibilism', (pp. 281 – 308) in Kane, R.H. (ed.) (2002) *The Oxford Handbook of Free Will* (First Edition), Oxford University Press
- Fischer, J.M. (2010) 'The Frankfurt Cases: The Moral of the Stories', *Philosophical Review*, Vol. 119, No. 3, pp. 315 – 336
- Fischer, J.M. (2011) 'Frankfurt-Type Examples and Semicompatibilism: New Work', (pp. 243 – 265) in Kane, R.H. (ed.) (2011) *The Oxford Handbook of Free Will* (Second Edition), Oxford University Press

- Fischer, J.M. & Ravizza, M. (1998) *Responsibility and Control: A Theory of Moral Responsibility*, Cambridge University Press
- Frankfurt, H. G. (1969) 'Alternate Possibilities and Moral Responsibility', *The Journal of Philosophy*, Vol. 66, No. 23. (Dec. 4, 1969), pp. 829 – 839
- Frankfurt, H.G. (1971) 'Freedom of the Will and the Concept of a Person', *The Journal of Philosophy*, Vol. 68, No. 1 (Jan. 14, 1971) pp. 5 – 20
- Frankfurt, H. G. (1988) *The Importance of What We Care About*, Cambridge University Press
- Ginet, C. (1980) 'The Conditional Analysis of Freedom', (pp. 171 – 186) in Van Inwagen (ed.) (1980) *Time and Cause*, Dordrecht.
- Ginet, C. (1996) 'In Defence of the Principle of Alternative Possibilities: Why I Don't Find Frankfurt's Argument Convincing', *Nous*, Vol. 30 – Supplement: Philosophical Perspectives, 10, *Metaphysics*, pp. 403 – 417
- Garnett, M. (2013) 'Fischer-style Compatibilism', *Analysis*, 73, (2), pp. 1 – 10
- Goetz, S. (2005) 'Frankfurt-Style Counterexamples and Begging the Question', *Midwest Studies in Philosophy*, Vol. XXIX, pp. 83 – 105
- Haji, I. (2009) *Incompatibilism's Allure: Principal Arguments for Incompatibilism*, Broadview Press.
- Harris, J.A. (2005) *Of Liberty and Necessity: The Free Will Debate in Eighteenth-Century British Philosophy*, Oxford University Press.
- Hunt, D. (2000) 'Moral Responsibility and Unavoidable Action,' *Philosophical Studies*, Vol. 97, No. 2, Jan 2000, pp. 195 – 227
- Hurley, S.L. (2000) 'Responsibility, Reason, and Irrelevant Alternatives', *Philosophy and Public Affairs*, Vol. 28, pp. 205 – 241
- Johnston, M. (1992) 'How to Speak of the Colours', *Philosophical Studies*, No. 68, pp. 221 – 263
- Kane, R.H. (1996) *The Significance of Free Will*, Oxford University Press.

- Kane, R.H. (ed.) (2002) *The Oxford Handbook of Free Will* (First Edition), Oxford University Press
- Kane, R.H. (ed.) (2011) *The Oxford Handbook of Free Will* (Second Edition), Oxford University Press
- Lehrer, K. (1968) 'Cans Without Ifs', *Analysis*, Vol. 29 (1. Oct, 1968), pp. 29 – 32 (reprinted in Watson, G. (ed.) (1982) *Free Will* (First Edition), OUP, pp. 41 – 45).
- Lewis, D. (1979) 'Counterfactual Dependence and Time's Arrow', *Nous*, Vol. 13, No. 4 (Nov. 1979), pp. 455 – 476
- Lewis, D. (1981) 'Are We Free to Break the Laws?', *Theoria*, No. 47 (1981), pp. 112 – 121.
- Lewis, D. (1987) *Philosophical Papers: Volume II*, Oxford University Press.
- Lewis, D. (1997) 'Finkish Dispositions', *The Philosophical Quarterly*, Vol. 47, No. 187, pp. 143 – 158
- Locke, J. (Woolhouse, R. (ed.). (1997) *An Essay Concerning Human Understanding*, Penguin Classics.
- Manley, D. & Wasserman, R. (2008) 'On Linking Dispositions and Conditionals', *Mind*, Vol. 117, No. 465, pp. 59 – 84
- Martin, C.B. (1994) 'Dispositions and Conditionals', *The Philosophical Quarterly*, Vol. 44, No. 174, pp. 1 – 8
- Maier, J. (2013) 'Abilities', *The Stanford Encyclopedia of Philosophy* (Summer 2013 Edition), Zalta, E.N. (ed.), URL = <<http://plato.stanford.edu/entries/abilities/>>
- McCann, H.J. (1998) *The Works of Agency: On Human Action, Will, and Freedom*. Cornell University Press. Ithaca.
- McKenna, M.S. (1997) 'Alternative Possibilities and the Failure of the Counterexample Strategy', *Journal of Social Philosophy*, Vol. 28, No. 3, Winter 1997, pp. 71 – 85
- McKenna, M.S. (1998) 'Does Strong Compatibilism Survive Frankfurt Counterexamples?', *Philosophical Studies*, Vol. 91, No. 3, pp. 259 – 264

- McKenna, M.S. (2011) 'Contemporary Compatibilism: Mesh Theories and Reason-Responsive Theories', (pp. 175 – 198) in Kane, R. (ed.) (2011) *The Oxford Handbook of Free Will (Second Edition)*, Oxford University Press
- McKittrick, J. (2003) 'A Case for Extrinsic Dispositions', *Australian Journal of Philosophy*, No. 81, pp. 155 – 174
- Mele, A.R. (2003) *Motivation and Agency*, Oxford University Press
- Mele, A.R. (2005) 'Libertarianism, Luck, and Control', *Pacific Philosophical Quarterly*, 86, pp. 381 – 407
- Mele, A.R. (2006) *Free Will and Luck*, Oxford University Press
- Mele, A.R. & Robb, D. (1998) 'Rescuing Frankfurt-style Cases', *The Philosophical Review*, Vol. 107, No. 1 (Jan. 1998), pp. 97 – 112
- Moore, G.E. (1911) *Ethics: and 'The Nature of Moral Philosophy'* (Shaw, W.H. (ed.)), Oxford University Press.
- Mumford, S. (1998) *Dispositions*, Oxford University Press.
- Nozick, R. (1981) *Philosophical Explanations*, Belknap Press, Harvard.
- O'Connor, T. (2011) 'Agent-Causal Theories of Freedom', (pp. 309 – 328) in Kane, R.H. (ed.) (2011) *The Oxford Handbook of Free Will (Second Edition)*, Oxford University Press
- Otsuka, M. (1998) 'Incompatibilism and the Avoidability of Blame', *Ethics*, vol. 108, No. 4, July 1998: pp. 685 – 701
- Pink, T. (2004) *Free Will: A Very Short Introduction*, Oxford University Press.
- Pereboom, D. (2007) 'Hard Incompatibilism', in *Four Views on Free Will – Great Debates in Philosophy* (Sosa, E. (series ed.)), Blackwell Publishing
- Ryle, G. (1959) *The Concept of Mind*, London.
- Sosa, E & Tooley, M (eds.) (1993) *Causation – Oxford Readings in Philosophy*, Oxford University Press

Shope, R.K. (1978) 'The Conditional Fallacy in Contemporary Philosophy', *The Journal of Philosophy*, Vol. LXXV, No. 8, pp. 397 – 413

Smith, M. (1997) 'A Theory of Freedom and Responsibility', in Cullity G. & Gaut, B. (eds.) *Ethics and Practical Reason*, Clarendon Press, Oxford.

Smith, M. (2003) 'Rational Capacities, or: How to Distinguish Recklessness, Weakness, and Compulsion', in Stroud, S. & Tappolet, C. (eds.) *Weakness of Will and Practical Irrationality*, Oxford University Press.

Strawson, P.F. (1962) 'Freedom and Resentment', *Proceedings of the British Academy*, vol. XLVIII, pp. 1 – 25

Stump, E. (1999) 'Alternative Possibilities and Moral Responsibility: The Flicker of Freedom', *The Journal of Ethics*, 3, pp. 299 – 324 (Kluwer Academic Publishers).

Torre, S. (2011) 'The Open Future', *Philosophy Compass*, 6/5, pp. 360 – 373

Van Inwagen, P. (1978) 'Ability and Responsibility', *The Philosophical Review*, Vol. 87, No. 2, Apr. 1978, pp. 201-224

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

Van Inwagen, P. (2002) 'Free Will Remains A Mystery' (pp. 158 – 177) in Kane, R.H. (ed.) (2002) *The Oxford Handbook of Free Will (First Edition)*, Oxford University Press

Vihvelin, K. (2000) 'Freedom, Foreknowledge, and the Principle of Alternate Possibilities', *Canadian Journal of Philosophy*, Vol. 30, No. 1 (Mar. 2000), pp. 1 – 23

Vihvelin, K. (2004) 'Free Will Demystified: A Dispositional Account', *Philosophical Topics*, Vol. 32, Nos. 1&2, pp. 427 – 450

Vihvelin, K. (2008) 'Foreknowledge, Frankfurt, and Ability to do Otherwise: A Reply to Fischer', *Canadian Journal of Philosophy*, Vol. 38, No. 3, pp. 343 – 372

Watson, G. (ed.) (1982) *Free Will – Oxford Readings in Philosophy (First Edition)*, Oxford University Press

Watson, G. (ed.) (2003) *Free Will – Oxford Readings in Philosophy* (Second Edition), Oxford University Press

Whittle, A. (2010) 'Dispositional Abilities', *Philosophers Imprint*, Vol. 10, No. 12 (Sept. 2010), pp. 1 – 23

Widerker, D. (1995) 'Libertarianism and Frankfurt's Attack on the Principle of Alternative Possibilities', *The Philosophical Review*, Vol. 104, No. 2, Apr. 1995, pp. 247-261

Widerker, D. & McKenna, M. (eds.) (2003) *Moral Responsibility and Alternative Possibilities: Essays on the Importance of Alternative Possibilities*, Ashgate, Aldershot.

Wolf, S. (1990) *Freedom Within Reason*, Oxford University Press.

Wyma, K.D. (1997) 'Moral Responsibility and Leeway For Action', *American Philosophical Quarterly*, Vol. 34, No. 1 (Jan. 1997), pp. 57 – 70