

*Living Without Free Will*

DERK PEREBOOM

*University of Vermont*



**CAMBRIDGE**  
UNIVERSITY PRESS

PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE  
The Pitt Building, Trumpington Street, Cambridge, United Kingdom

CAMBRIDGE UNIVERSITY PRESS  
The Edinburgh Building, Cambridge CB2 2RU, UK  
40 West 20th Street, New York, NY 10011-4211, USA  
10 Stamford Road, Oakleigh, VIC 3166, Australia  
Ruiz de Alarcón 13, 28014 Madrid, Spain  
Dock House, The Waterfront, Cape Town 8001, South Africa  
<http://www.cambridge.org>

© Derk Pereboom 2001

This book is in copyright. Subject to statutory exception  
and to the provisions of relevant collective licensing agreements,  
no reproduction of any part may take place without  
the written permission of Cambridge University Press.

First published 2001

Printed in the United States of America

*Typeface* Bembo 10.5/13 pt.      *System* QuarkXPress [BTS]

*A catalog record for this book is available from the British Library.*

*Library of Congress Cataloging in Publication Data*

- Pereboom, Derk, 1957–  
Living without free will / Derk Pereboom.  
p. cm. – (Cambridge studies in philosophy)  
Includes bibliographical references.  
ISBN 0-521-79198-7 (hardcover)
1. Free will and determinism. I. Title. II. Series.  
BJ1461 .P47 2001  
123'.5 – dc21                      00-059877

ISBN 0 521 79198 7 hardback

# Contents

Acknowledgments	<i>page</i> xi
Introduction: Hard incompatibilism	xiii
1 Alternative possibilities and causal histories	1
Two incompatibilist intuitions	1
Leeway vs. causal-history incompatibilism	2
Flickers and robustness	6
A libertarian objection to Frankfurt-style arguments	8
Problems for recent attempts to answer Widerker	10
Frankfurt-style scenarios without prior signs	14
A new Frankfurt-style scenario	18
A different kind of alternative-possibility condition	23
Ginet's challenge	28
An objection to causal-history incompatibilism	33
A final word	36
2 Coherence objections to libertarianism	38
Introduction	38
The Humean challenge to event-causal libertarianism	41
Randomizing manipulators	50
Why the Humean challenge fails to undermine agent-causal libertarianism	55
Objections to the agent-causal theory's notion of causation	59
A disparity between causation and explanation?	63
Galen Strawson's argument against the coherence of libertarianism	65
The prospects for agent-causal libertarianism	68
3 Empirical objections to agent-causal libertarianism	69
Introduction	69
Problems for the ordinary nonreductivist strategy	70

Strong emergentism and microphysical constraints	73
Strong emergentism, non-physicalism, and the prospects for reconciliation with physics	79
Difficulties for the overriding approach	85
Why invoking <i>ceteris paribus</i> laws fails to advance the libertarian's case	86
Libertarianism and common sense	88
4 Problems for compatibilism	89
Challenging compatibilism	89
The first route to compatibilism: Determinism as irrelevant to responsibility	90
Criticisms of Strawson's view	94
The second route to compatibilism: Causal integrationist conditions	100
Why these four compatibilist accounts can be undermined	110
Haji's response to an earlier version of the four-case argument	117
Fischer and Ravizza's proposed sufficient condition	120
Strawson and Wallace on generalization strategies	123
Conclusion	125
5 The contours of hard incompatibilism	127
Hard incompatibilism defined and defended	127
Contrasts with similar views	128
Agency	135
Dispensing with blameworthiness and praiseworthiness	139
Does hard incompatibilism undermine morality?	141
Why hard incompatibilism is not restricted to consequentialism in ethics	148
Moral worth	152
Wrongdoing	154
6 Hard incompatibilism and criminal behavior	158
Is there an acceptable hard incompatibilist position on managing criminal behavior?	158
Hard incompatibilism rules out retributivism	159
Hard incompatibilism and the moral education theory	161
Hard incompatibilism and deterrence theories: The utilitarian version	166
Deterrence justified by the right to self-defense	168
Problems for the self-defense theory	172
Quarantine and criminal detention	174
Rehabilitation	178
Reviewing the conclusions	186

7	Hard incompatibilism and meaning in life	187
	Introduction	187
	Life projects: Sartre and Honderich	188
	Affirmation	191
	Can belief in free will be justified pragmatically?	197
	Reactive attitudes and interpersonal relationships	199
	Guilt and repentance	204
	The advantages of hard incompatibilism	207
	Bibliography	215
	Index	225

# 1

## *Alternative Possibilities and Causal Histories*

### TWO INCOMPATIBILIST INTUITIONS

The claim that moral responsibility for an action requires that the agent could have done otherwise is surely attractive. Moreover, it seems reasonable to contend that a requirement of this sort is not merely a necessary condition of little consequence, but that it plays a significant role in explaining why an agent is morally responsible. For if an agent is to be blameworthy for an action, it seems crucial that she could have done something to avoid being blameworthy – that she could have done something to get herself off the hook. If she is to be praiseworthy for an action, it seems important that she could have done something less admirable. Libertarians have often grounded their incompatibilism precisely in such intuitions. As a result, they have often defended the following principle of alternative possibilities:

- (1) An action is free in the sense required for moral responsibility only if the agent could have done otherwise than she actually did.

or a similar principle about choice:

- (2) An action is free in the sense required for moral responsibility only if the agent could have chosen otherwise than she actually did.

I shall argue that despite resourceful attempts to defend conditions of this sort, any such requirement that is relevant to explaining why an agent is morally responsible for an action falls to counterexamples. I maintain instead that the most plausible and fundamentally explanatory incompatibilist principles concern the causal history of an action, and

not alternative possibilities.<sup>1</sup> These claims leave open the prospect of alternative-possibilities conditions necessary for moral responsibility but nevertheless irrelevant to explaining why an agent is morally responsible. I believe that there could well be such conditions.

#### LEEWAY VS. CAUSAL HISTORY INCOMPATIBILISM

Familiarly, arguments of the kind devised by Harry Frankfurt provide an especially formidable challenge to alternative possibility conditions.<sup>2</sup> The standard versions deploy examples with a particular sort of structure. Here is one of Fischer's cases:

Black is a nefarious neurosurgeon. In performing an operation on Jones to remove a brain tumor, Black inserts a mechanism into Jones's brain which enables Black to monitor and control Jones's activities. Jones, meanwhile, knows nothing of this. Black exercises this control through a computer which he has programmed so that, among other things, it monitors Jones's voting behavior. If Jones shows an inclination to decide to vote for Carter, then the computer, through the mechanism in Jones's brain, intervenes to assure that he actually decides to vote for Reagan and does so vote. But if Jones decides on his own to vote for Reagan, the computer does nothing but continue to monitor – without affecting the goings-on in Jones's head. Suppose Jones decides to vote for Reagan on his own, just as he would have if Black had *not* inserted the mechanism into his head.<sup>3</sup>

1. I argued for this view in "Determinism *Al Dente* (1985), and later in "Alternative Possibilities and Causal Histories," *Philosophical Perspectives* 14 (2000). For similar positions, see Eleanore Stump, "Intellect, Will, and the Principle of Alternate Possibilities," in *Christian Theism and the Problems of Philosophy*, ed. Michael Beatty (Notre Dame, IN: University of Notre Dame Press, 1990), pp. 254–85, reprinted in *Moral Responsibility*, ed. John Martin Fischer and Mark Ravizza (Ithaca: Cornell University Press, 1993), pp. 237–62; "Libertarian Freedom and the Principle of Alternative Possibilities," in *Faith, Freedom, and Rationality*, ed. Jeff Jordan and Daniel Howard Snyder (Lanham, MD: Rowman and Littlefield, 1996), pp. 73–88; Linda Zagebski, *The Dilemma of Freedom and Foreknowledge* (New York: Oxford University Press, 1991), Chapter 6, Section 2.1; "Does Libertarian Freedom Require Alternate Possibilities?" *Philosophical Perspectives* 14 (2000); Robert Heinaman, "Incompatibilism without the Principle of Alternative Possibilities," *Australasian Journal of Philosophy* 64 (1986), pp. 266–76; Michael Della Rocca, "Frankfurt, Fischer, and Flickers," *Noûs* 32 (1998), pp. 99–105; David Hunt, "Moral Responsibility and Unavoidable Action," *Philosophical Studies* 97 (2000), pp. 195–227.
2. Harry G. Frankfurt, "Alternate Possibilities and Moral Responsibility," *Journal of Philosophy* 1969, pp. 829–839; John Martin Fischer, "Responsibility and Control," in *Moral Responsibility*, Fischer, ed. (Ithaca: Cornell University Press, 1986), pp. 174–190.
3. Fischer, "Responsibility and Control," p. 176.

Fischer's intuition is that Jones could be responsible for voting or deciding to vote for Reagan, although he could not have done or chosen otherwise. Jones could not have done or even have chosen otherwise, because the device would have arrested the deliberative process before it resulted in any alternative choice. The conclusion of the argument is that conditions (1) and (2) are mistaken.

Fischer has contended that this type of argument does not refute the claim that moral responsibility requires that the actual causal history of the action not be deterministic. It leaves untouched the view that moral responsibility requires that one's action not actually result from a deterministic causal process that traces back to factors beyond one's control – back to causal factors that one could not have produced, altered, or prevented.<sup>4</sup> I believe that this contention of Fischer's is correct. Notice that this Frankfurt-style case does not specify that Jones's action is causally determined in this way. If it were specified that his choice is deterministically produced by factors beyond his control, by, for example, Martian neuroscientists, then the intuition that he could be morally responsible might well fade away. Furthermore, it seems possible for one's action not to result from a deterministic causal process that traces back to factors beyond one's control while one cannot do or choose otherwise. For, as is clear from the Frankfurt-style case, the factors that make it so that an agent cannot do or choose otherwise need not also determine him to act as he does, since they need not be part of the actual causal history of his action at all.

This reflection suggests a different requirement on the sort of freedom we are seeking to characterize:

- (3) An action is free in the sense required for moral responsibility only if it is not produced by a deterministic process that traces back to causal factors beyond the agent's control.

Condition (3) specifies a necessary condition on the sort of freedom required for moral responsibility that I believe any incompatibilist should endorse. One might note that even if it is not a necessary condition on moral responsibility that the agent could have done or chosen otherwise, the incompatibilist can still claim that one is not morally responsible for an action if one could not have done or chosen otherwise due to the choice's resulting from a deterministic causal process that traces back to factors beyond one's control.

4. Fischer, "Responsibility and Control," pp. 182–85.



In his central condition on moral responsibility (UR, for “ultimate responsibility”), Kane expresses one aspect of this intuition very nicely:

(U): For every X and Y (where X and Y represent occurrences of events and/or states), if the agent is personally responsible for X, and if Y is an *arche* (or sufficient ground or cause or explanation) for X, then the agent must also be personally responsible for Y.

(Kane spells out the alternative-possibilities intuition in the (R)-part of (UR).)<sup>5</sup> Conditions such as (3) and (U), I believe, have a critical role in explaining why agents would be morally responsible. If such conditions are not met by an agent’s decision, he lacks a certain kind of control over this decision, and it is for this reason that he is not morally responsible. The sort of control at issue is that the agent must in an appropriate sense be the ultimate source or cause of the action. Kane expresses the point in this way:

What (U) thus requires is that if an agent is ultimately responsible for an action, the action cannot have a sufficient reason of any of these kinds *for which the agent is not also responsible*. If the action did have such a sufficient reason for which the agent was not responsible, then the action, or the agent’s will to perform it, would have its source in something the agent played no role in producing. Then the *arche* of the action, or of the agent’s will to perform it, would not be “in the agent,” but in something else.<sup>6</sup>

What lies at the core of the intuition expressed by (3) and (U) is a claim about origination, which might be formulated as follows:

(O) If an agent is morally responsible for her deciding to perform an action, then the production of this decision must be something over which the agent has control, and an agent is not morally responsible for the decision if it is produced by a source over which she has no control.

5. Robert Kane, *The Significance of Free Will* (New York: Oxford University Press, 1996), p. 35. (UR) in its entirety is: (UR): An agent is *ultimately responsible* for some (event or state) E’s occurring only if (R) the agent is personally responsible for E’s occurring in a sense which entails that something the agent voluntarily (or willingly) did or omitted, and for which the agent could have voluntarily done otherwise, either was, or causally contributed to, E’s occurrence and made a difference to whether or not E occurred; and (U) for every X and Y (where X and Y represent occurrences of events and/or states) if the agent is personally responsible for X, and if Y is an *arche* (or sufficient ground or cause or explanation) for X, then the agent must also be personally responsible for Y. Thus, if there is a sufficient ground for an agent’s decision in events that precede the agent’s birth (together with laws of nature), then presuming that an agent cannot be personally responsible for events that precede her birth or for laws of nature, she cannot be personally responsible for the decision.
6. Kane, *The Significance of Free Will*, p. 73.

Ted Honderich also stresses the importance of a notion of origination for our sense of moral responsibility.<sup>7</sup> I think that (O) expresses the most fundamental and plausible incompatibilist intuition about how an agent's moral responsibility is grounded.<sup>8</sup> It explains not only why one might think that determinism and moral responsibility are incompatible, but also why one might believe that an agent cannot be morally responsible for a decision if it occurs without any cause whatsoever. For such a decision is produced by nothing, and hence the production of the decision is not something over which the agent has control. I shall clarify this condition and examine the surrounding issues more thoroughly in Chapter 2.

We might call those who incline toward the view that an alternative possibilities condition has the more important role in explaining why an agent would be morally responsible *leeway incompatibilists*, and those who are predisposed to maintain that an incompatibilist condition on the causal history of the action plays the more significant part *causal history incompatibilists*.<sup>9</sup> Leeway incompatibilists would argue that the actual causal history of a morally responsible action must be

7. Ted Honderich, *A Theory of Determinism* (Oxford: Oxford University Press, 1988), e.g., pp. 194–206.
8. Gary Watson, although he is not an incompatibilist, also maintains that the condition on origination is the fundamental incompatibilist claim; “Responsibility and the Limits of Evil,” in *Responsibility, Character and the Emotions*, ed. Ferdinand Schoeman (Cambridge: Cambridge University Press, 1987), pp. 256–86, at p. 282.
9. In “Compatibilists Could Have Done Otherwise: Responsibility and Negative Agency” (*The Philosophical Review* 103 (1994), pp. 453–88), Alison McIntyre convincingly argues that an analog of Frankfurt’s argument undermines the Principle of Possible Actions for omissions. But she also attempts to undermine any indeterminist requirement for moral responsibility with a Frankfurt–style case (pp. 472–78). A princess rises from her seat at the opera for a photo opportunity, and while it is customary for her to sit down after one minute, she decides to stand for four minutes. But a scientist has placed a force field around her, so that had she decided to sit down after one minute she would have remained standing for an additional three. Yet it is clear she is morally responsible for standing for the four minutes. About this case McIntyre says: “. . . even if her decision to stand for four minutes is not causally determined, it is nevertheless causally determined, once she has stood for a minute, that she will stand for three more minutes. To grant that the Princess can be morally responsible for standing for the last three minutes is *ipso facto* to grant that an agent can be morally responsible for behavior that is causally determined.” But this is not a situation in which an agent is responsible for an action that is produced by a deterministic process that traces back to factors beyond her control. McIntyre’s case specifies external factors that prevent the Princess from performing the action in question, but these external factors play no role in the actual causal history of the action. In fact, her case is consistent with the action’s being freely produced by a libertarian agent–causal power. Hence, condition (3) survives McIntyre’s argument, and I maintain that this condition withstands any argument that employs a Frankfurt–style strategy.

indeterministic, but they would be amenable to the claim that this is so only because an indeterministic history is required to secure alternative possibilities. Causal history incompatibilists would lean toward the position that the role the causal history plays in explaining why an agent would be morally responsible is independent of facts about alternative possibilities.

Against causal history incompatibilism, Fischer argues that “there is simply no good reason to suppose that causal determinism in itself (and apart from considerations pertaining to alternative possibilities) vitiates our moral responsibility.”<sup>10</sup> Fischer, I believe, is mistaken on this point. To be sure, one incompatibilist intuition that we seem naturally to have is that if we could in no sense do otherwise, then we could never have refrained from the wrongful actions we perform, and thus we cannot legitimately be held blameworthy for them. But another very powerful and common intuition is that if all of our behavior were “in the cards” before we were born – in the sense that things happened before we came to exist that, by way of a deterministic causal process, inevitably result in our behavior – then we cannot legitimately be blamed for our wrongdoing. By this intuition, if causal factors existed before a criminal was born that by way of a deterministic process, inevitably issue in his act of murder, then he cannot legitimately be blamed for his action. If all of our actions had this type of causal history, then it would seem that we lack the kind of control over our actions that moral responsibility requires.

Now I do not believe that in the dialectic of the debate, one should expect Fischer, or any compatibilist, to be moved much by this incompatibilist intuition alone to abandon his position. In my view, the more powerful, and indeed the best, type of challenge to compatibilism develops the claim that causal determination presents in principle no less of a threat to moral responsibility than does covert manipulation. We shall turn to that challenge in Chapter 4. Nonetheless, what this intuition should show at this stage is that there might well be a coherent incompatibilist position that could survive the demise of alternative-possibilities requirements.<sup>11</sup>

#### FLICKERS AND ROBUSTNESS

Thus in my view it is the intuition expressed by (O) rather than one associated with an alternative-possibility condition that is the most fun-

10. Fischer, *The Metaphysics of Free Will* (Oxford: Blackwell, 1994), p. 159.

11. Della Rocca, in “Frankfurt, Fischer, and Flickers,” develops a similar theme.

damental and plausible underlying ground for incompatibilism. But this claim has not yet been thoroughly tested. Perhaps some version of an alternative-possibilities condition on moral responsibility can survive any Frankfurt-style argument. Libertarians have contended that according to any argument of this kind, there must be some factor that the neurophysiologist's device is rigged up to detect that could but does not actually occur in the agent, such as an intention to do otherwise.<sup>12</sup> The possible occurrence of this factor – this “flicker of freedom,” to use Fischer's term – might then function as the alternative possibility that is required for moral responsibility.<sup>13</sup> Libertarians, in particular, are predisposed to locate the source of moral responsibility in the will, and if moral responsibility requires alternative possibilities, it must require, more precisely, the possibility of willing to do otherwise. But it is not implausible that the formation of an intention to do otherwise should count as willing to do otherwise, and hence the possibility of forming such an intention would assist in explaining moral responsibility for the choice or action at issue.

Fischer, however, argues that one can construct different Frankfurt-style stories in which the intervening device detects some factor prior to the formation of the intention. One might, for example, imagine that Jones will decide to kill Smith only if Jones blushes beforehand. Then Jones's failure to blush (by a certain time) might be the alternative possibility that would trigger the intervention that causes him to kill Smith. Supposing that Jones acts without intervention, we might well have the intuition that he is morally responsible, despite the fact that he could not have done or chosen otherwise, or formed an alternative intention. He could have failed to blush, but as Fischer argues, such a flicker is of no use to the libertarian, since it is not sufficiently *robust*, it is too “flimsy and exiguous” to play a part in grounding moral responsibility.<sup>14</sup>

I agree with Fischer, and here is a first pass at characterizing robustness. The intuition underlying the alternative-possibilities requirement

12. Peter van Inwagen, *An Essay on Free Will* (Oxford: Oxford University Press, 1983), pp. 166–80. In my view, the intention to perform an action is produced by the choice to perform the action, and hence succeeds and does not precede it. Thus an intention to perform an action could not serve as a sign for intervention that would preclude a choice to perform it.
13. Fischer provides a lucid discussion and criticism of this strategy in *The Metaphysics of Free Will*, pp. 134–47.
14. Fischer, *The Metaphysics of Free Will*, pp. 131–59; “Recent Work on Moral Responsibility,” *Ethics* 110 (1999), pp. 93–139.

is that if, for example, an agent is to be blameworthy for an action, it is crucial that he could have done something to avoid being blameworthy. If having an alternative possibility does in fact play a role in explaining an agent's moral responsibility for an action, it would have to be robust at least in the sense that as a result of securing that alternative possibility, the agent would *thereby* have avoided the responsibility he has for the action he performed – it would be his securing of that alternative possibility per se that would explain why the agent would have avoided the responsibility. Failing to blush in Fischer's scenario does not meet this criterion of robustness. For if Jones had failed to blush, he would not thereby have avoided responsibility for evading killing Smith – it would not be the failure to blush per se that would explain why Jones would not be blameworthy. By typical libertarian intuitions, one robust sort of alternative possibility would involve willing to do otherwise than to perform the action the agent in fact wills to perform.<sup>15</sup>

#### A LIBERTARIAN OBJECTION TO FRANKFURT-STYLE ARGUMENTS

It might now seem that any alternative-possibilities condition on moral responsibility can be defeated by a Frankfurt-style argument that employs a non-robust flicker of freedom. But perhaps this line of defense for Frankfurt-style arguments is too quick. An important kind of objection against these sorts of arguments was initially raised by Kane and then systematically developed by David Widerker. (A close relative has been advanced by Carl Ginet, which we will consider shortly.<sup>16</sup>) The general form of the Kane/Widerker objection is this. For any Frankfurt-style case, if causal determinism is assumed, the libertarian will not have, and cannot be expected to have, the intuition that the agent is morally responsible. If, on the other hand, libertarian indeter-

15. See also Mele's characterization of robustness, which I endorse, in "Soft Libertarianism and Frankfurt-Style Scenarios," *Philosophical Topics* 24 (1996), pp. 123–41, at pp. 126–7.
16. Kane, *Free Will and Values* (Albany: SUNY Press, 1985), p. 51 n. 25, and *The Significance of Free Will*, pp. 142–4, 191–2; David Widerker, "Libertarianism and Frankfurt's Attack on the Principle of Alternative Possibilities," *The Philosophical Review* 104 (1995), pp. 247–61; Carl Ginet, "In Defense of the Principle of Alternative Possibilities: Why I Don't Find Frankfurt's Arguments Convincing," *Philosophical Perspectives* 10 (1996), pp. 403–17; see also Keith D. Wyma, "Moral Responsibility and Leeway for Action," *American Philosophical Quarterly* 34 (1997), pp. 57–70. Fischer provides a clear and helpful account of these views in "Recent Work on Moral Responsibility," pp. 111–12.

minism is presupposed, an effective Frankfurt-style scenario cannot be devised, for any such case will fall to a dilemma. In Frankfurt-style cases, the actual situation always features a prior sign by which the intervener can know that the agent will perform the action he does, and that signals the fact that intervention is not necessary. If in the proposed case, the sign causally determines the action, or if it is associated with something that does so, the intervener's predictive ability can be explained. But then the libertarian would not have the intuition that the agent is morally responsible. If the relationship between the sign and the action is not causally deterministic in such ways, then the libertarian can claim that the agent could have done otherwise despite the occurrence of the prior sign. Either way, some principle of alternative possibilities emerges unscathed.

Widerker's particular version of the objection has the following structure.<sup>17</sup> The case at issue is the one we have just encountered, in which Jones wants to kill Smith, but Black is afraid that Jones might become fainthearted, and so he is prepared to intervene if Jones fails to show a sign that he will kill Smith. The sign that he will kill Smith is his blushing at  $t_1$ . The important features of the scenario are these:

- (1) If Jones is blushing at  $t_1$ , then, provided no one intervenes, he will decide at  $t_2$  to kill Smith.
- (2) If Jones is not blushing at  $t_1$ , then, provided no one intervenes, he will not decide at  $t_2$  to kill Smith.
- (3) If Black sees that Jones shows signs that he will not decide at  $t_2$  to kill Smith – that is, he sees that Jones is *not* blushing at  $t_1$  – then Black will force Jones to decide at  $t_2$  to kill Smith; but if he sees that Jones is blushing at  $t_1$ , then he will do nothing.

Finally, suppose that Black does not have to show his hand, because

- (4) Jones is blushing at  $t_1$ , and decides at  $t_2$  to kill Smith for reasons of his own.<sup>18</sup>

Although the case is meant to show that Jones is morally responsible despite the fact that he could not have done otherwise, Widerker claims that this conclusion is not forced on the libertarian:

17. Cf. Ishtayaque Haji, *Moral Appraisability* (New York: Oxford University Press, 1998), pp. 34–5.  
18. Widerker, "Libertarianism and Frankfurt's Attack on the Principle of Alternative Possibilities," pp. 249–50.

Note that the truth of (1) cannot be grounded in the fact that Jones's blushing at  $t_1$  is, in the circumstances, causally sufficient for his decision to kill Smith, or in the fact that it is indicative of a state that is causally sufficient for that decision, since such an assumption would . . . [not be] accepted by the libertarian. On the other hand, if (1) is not thus grounded, then the following two options are available to the libertarian to resist the contention that Jones's decision to kill Smith is unavoidable. He may either reject (1), claiming that the most that he would be prepared to allow is

(1a) If Jones is blushing at  $t_1$ , then Jones will *probably* decide at  $t_2$  to kill Smith . . .

But (1a) is compatible with Jones's having the power to decide not to kill Smith, since there remains the possibility of Jones's acting out of character. Or the libertarian may construe (1) as a conditional of freedom in Plantinga's sense . . . that is, as

(1b) If Jones is blushing at  $t_1$ , then Jones will *freely* decide at  $t_2$  to kill Smith, [in a sense that allows that the agent could have decided otherwise]<sup>19</sup>

in which case the libertarian may again claim that in the actual situation when Jones is blushing at  $t_1$ , it is within his power to refrain from deciding to kill Smith at  $t_2$ .<sup>20</sup>

Widerker's is a very important objection, and it serves as a test for the effectiveness of any Frankfurt-style argument. One point of clarification: If the libertarian that Widerker supposes Frankfurt must convince is simply presupposing a principle of alternative possibilities, then one could not expect that a Frankfurt-style argument would dislodge his view. But Widerker, I think, does not intend that his libertarian simply presuppose this principle, but rather only the claim that moral responsibility is incompatible with an action's having a deterministic causal history. I will proceed with this understanding of Widerker's objection.

#### PROBLEMS FOR RECENT ATTEMPTS TO ANSWER WIDERKER

Several critics have tried to construct Frankfurt-style arguments that escape this objection. The cases used in these arguments divide into two categories:

19. This bracketed phrase does not occur in Widerker's text, but it clearly expresses his meaning.
20. Widerker, "Libertarianism and Frankfurt's Attack on the Principle of Alternative Possibilities," p. 250.

(a) Those in which the relationship between the prior sign and the action is causally deterministic, and the indeterminism that makes for the agent's libertarian freedom is present in the causal history of the action before the prior sign,

and

(b) those in which the prior sign is eliminated altogether.

Eleonore Stump and Ishtiyaque Haji have constructed examples in category (a),<sup>21</sup> while David Hunt and Alfred Mele, together with David Robb, have devised scenarios in category (b).<sup>22</sup>

In my view, the cases that have been devised in each of these categories face significant problems. First, (a)-type situations are difficult to construct so that they are effective against Widerker's objection. Stump's and Haji's examples have serious drawbacks. In Stump's case, Grey, the neurosurgeon, wants to ensure that Jones will vote for Reagan. Grey finds that every time Jones decides to vote for Republicans, the decision regularly correlates with the completion of a sequence of neural firings in Jones's brain that always includes, near the beginning, the firing of neurons a, b, and c. Jones's deciding to vote for Democratic candidates is correlated with the completion of a neural sequence that always includes, near the beginning, the firing of neurons x, y, and z. Whenever Grey's neuroscope detects the firing of x, y, and z, it disrupts that sequence, with the result that the sequence is not brought to completion. Instead, the device activates a coercive mechanism that makes Jones vote Republican. Crucially, Stump specifies that the firing of x, y, and z does not constitute a decision, and in her view the occurrence of this sequence would not count as a robust alternative possibility. If, on the other hand, the neuroscope detects the firing of a, b, and c, it allows the sequence to proceed to completion and the decision to vote Republican to occur.<sup>23</sup> Stump specifies that the decision is indeed a causal outcome of the neural sequence.<sup>24</sup> What makes the agent libertarian is that the neural sequence is not the outcome of a causal chain that originates in a cause outside him. Rather, it is the outcome of a causal chain that originates, at least to a significant extent, in an act of

21. Stump, "Libertarian Freedom and the Principle of Alternative Possibilities"; Haji, *Moral Appraisability*, p. 36.

22. Alfred Mele and David Robb, "Rescuing Frankfurt-Style Cases," *The Philosophical Review* 107 (1998), pp. 97–112; David Hunt, "Moral Responsibility and Unavoidable Action."

23. Stump, "Libertarian Freedom and the Principle of Alternative Possibilities," pp. 77–8.

24. *Ibid.*, p. 79.



the agent which is not the outcome of a causal chain that originates in a cause outside the agent. Here, Stump suggests the Aquinas-inspired view that the neural sequence is the outcome of a causal chain that originates in the agent's intellect and will.<sup>25</sup>

But as Stewart Goetz points out, to assess this case, one needs to know more about the psychological features of the act performed by the agent to cause the neural process. If this originating act is causally determined, then Stump's agent would appear not to be free in the libertarian sense. If it is not causally determined, then he might well have robust alternative possibilities for action. If the originating act is an intention to make a decision, for example, and if the indeterminism of that act allows for the agent to have avoided intending to make the decision, then the case might well include a robust alternative possibility after all.<sup>26</sup> Note that in Stump's setup, the agent's performance of that act – which constitutes the agent's crucial libertarian causal role – precedes the possible intervention.<sup>27</sup>

More generally, the challenge for Stump is to characterize the agent's causal role so that (i) her action is not causally determined (by factors beyond her control), and (ii) her action does not involve robust alternative possibilities. A case of the sort that Stump devises is subject to the following dilemma: If the indeterminism (whether or not it is a characteristic of the sort of agent's act she has in mind) that occurs prior to the neural sequence is significant enough to make the action a libertarian freely willed action, then it has not been ruled out that the indeterministic juncture features a robust alternative possibility. If Stump were to reject the claim that there is a robust alternative possibility at this point, then it would remain open to a libertarian (like Widerker) to deny that the agent has genuine libertarian free will. Perhaps it is possible to embellish Stump's example to answer this objection. However, it is not clear to me that there could be a plausible Frankfurt-style case in which the action is not causally determined by factors beyond the agent's control (in a way that would satisfy the libertarian) and she lacks robust alternative possibilities if the intervention would occur after the crucial indeterministic juncture.

25. Stump, "Libertarian Freedom and the Principle of Alternative Possibilities," pp. 80–5.

26. Stewart Goetz, "Stumping for Widerker," *Faith and Philosophy* 16 (1999), pp. 83–9. In this article, he develops this and other criticisms in further detail.

27. Stump replies to Goetz's objection in "Dust, Determinism and Frankfurt: A Reply to Goetz," in *Faith and Philosophy* 16 (1999), pp. 413–22, but in my view she does not lay to rest the worry I just described.

In Haji's example, the sort of libertarian agency attributed to the agent consists in its being undetermined which of various considerations will enter the mind of the agent in deliberation. So, at the outset of Jones's deliberation, it is causally open whether he will kill Smith, because it is causally undetermined whether various considerations will enter his mind at the onset of his musings. The infallible predictor, Black,

intervenes if and only if he believes that Jones will not make the decisive best judgment that favors the decision to kill Smith which he, Black, wants Jones to make. Specifically, should Jones make the judgment that he ought not to kill Smith, *then* (and only then) will Black intervene and cause Jones to alter the judgment.<sup>28</sup>

There is no need for Black to intervene "as Jones decides appropriately on his own," and one will have the intuition that Jones could be morally responsible for his decision. One problem for Haji's case is that it was open to Jones to have made the decisive best judgment that he ought not to kill Smith, and this alternative possibility seems robust. For it appears plausible that if Jones had made the decisive best judgment that he ought not to kill Smith, he would thereby have avoided the responsibility he has for the action he actually performed – it would be his securing of this alternative possibility per se that would explain why he would have avoided this responsibility. Another difficulty for Haji's example is that many libertarians would not let the sort of indeterminacy he specifies be significant for moral responsibility because it fails to provide the agent with enhanced control. A case in which the relevant considerations indeterministically enter the mind of the agent, whereupon his judgment and decision are determined, would seem to exhibit no more control by him than a situation in which such considerations deterministically enter the mind of the agent, whereupon the agent's judgment and decision are causally determined.<sup>29</sup>

28. Haji, *Moral Appraisability*, p. 36. Haji attributes the inspiration for this sort of libertarianism to Mele, in *Autonomous Agents* (New York: Oxford University Press, 1995), p. 216. This kind of libertarianism is also suggested by Daniel Dennett, in "Giving Libertarians What They Say They Want," in his *Brainstorms* (Montgomery, VT: Bradford Books, 1978), at p. 295.
29. See, for example, Randolph Clarke, "Agent Causation and Event Causation in the Production of Free Action," *Free Will*, ed. Derk Pereboom (Indianapolis: Hackett, 1997), pp. 273–300, at p. 286 (the original and longer version of this article appears in *Philosophical Topics* 24, 1996, pp. 19–48).

## FRANKFURT-STYLE SCENARIOS WITHOUT PRIOR SIGNS

Cases in category (b) exemplify a different strategy for opposing alternative-possibility conditions. In these cases, there are no prior signs to guide intervention, not even non-robust flickers of freedom. One ingenious scenario in this category is presented by Mele and Robb.<sup>30</sup> The example features Bob, who inhabits a world in which determinism is false:

At  $t_1$ , Black initiates a certain deterministic process  $P$  in Bob's brain with the intention of thereby causing Bob to decide at  $t_2$  (an hour later, say) to steal Ann's car. The process, which is screened off from Bob's consciousness, will deterministically culminate in Bob's deciding at  $t_2$  to steal Ann's car unless he decides on his own to steal it or is incapable at  $t_2$  of making a decision (because, e.g., he is dead at  $t_2$ ). . . . The process is in no way sensitive to any "sign" of what Bob will decide. As it happens, at  $t_2$  Bob decides on his own to steal the car, on the basis of his own indeterministic deliberation about whether to steal it, and his decision has no deterministic cause. But if he had not just then decided on his own to steal it,  $P$  would have deterministically issued, at  $t_2$ , in his deciding to steal it. Rest assured that  $P$  in no way influences the indeterministic decision-making process that actually issues in Bob's decision.

Mele and Robb claim that Bob is plausibly morally responsible for his decision. I think that their argument may in fact be successful, but that their development of the case raises one problem that could undermine it. Mele and Robb discuss several potential problems for their scenario, one of which is whether we can make sense of what would happen at  $t_2$  if  $P$  and Bob's indeterministic deliberative process were to diverge at  $t_2$ . Here is how they handle the difficulty:

The issue may be pictured, fancifully, as follows.<sup>31</sup> Two different "decision nodes" in Bob's brain are directly relevant. The "lighting up" of node  $N_1$  represents his deciding to steal the car, and the "lighting up" of node  $N_2$  represents his deciding *not* to steal the car. Under normal circumstances and in the absence of preemption, a process's "hitting" a decision node in Bob "lights up" that node. If it were to be the case both that  $P$  hits  $N_1$  at  $t_2$  and that  $x$  does not hit  $N_1$  at  $t_2$ , then  $P$  would light up  $N_1$ . If both processes were to hit  $N_1$

30. Mele and Robb, "Rescuing Frankfurt-Style Cases."

31. (Mele and Robb's note.) The picture obviously is neuro-fictional, but it is still useful nonetheless.

at  $t_2$ , Bob's indeterministic deliberative process,  $x$ , would light up N1 and P would not. The present question is this. What would happen if, at  $t_2$ , P were to hit N1 and  $x$  were to hit N2? That is, what would happen if the two processes were to "diverge" in this way? And *why*?

We extend Bob's story as follows. Although if both processes were to hit N1 at  $t_2$ , Bob's indeterministic deliberative process,  $x$ , would preempt P and light up N1, it is also the case that if, at  $t_2$ , P were to hit N1 and  $x$  were to hit N2, P would prevail. In the latter case, P would light up N1 and the indeterministic process would not light up N2. Of course, readers would like a story about why it is that although  $x$  would preempt P in the former situation, P would prevail over  $x$  in the latter. Here is one story. By  $t_2$ , P has "neutralized" N2 (but without affecting what goes on in  $x$ ). That is why, if  $x$  were to hit N2 at  $t_2$ , N2 would not light up.<sup>32</sup> More fully, by  $t_2$  P has neutralized all of the nodes in Bob for decisions that are contrary to a decision at  $t_2$  to steal Ann's car (e.g., a decision at  $t_2$  not to steal anyone's car and a decision at  $t_2$  never to steal anything).<sup>33</sup> In convenient shorthand, by  $t_2$  P has neutralized N2 and all its "cognate decision nodes." Bear in mind that all we need is a conceptually possible scenario, and this certainly looks like one.<sup>34</sup>

The aspect of this story that might raise the libertarian's eyebrows is P's neutralization of N2 and all its cognate decision nodes. For he might be tempted to claim that P's neutralizing procedure is equivalent to P's causal determination of Bob's decision to steal the car. On the other hand, Mele and Robb do specify that P's neutralizing activity does not affect what goes on in Bob's indeterministic decision-making process, and so it would seem that P would not causally determine the decision. How can we shed light on this difficulty?

Let us examine an approach in category (b) that more vigorously exploits the neutralization idea. A strategy of this type has become known as "blockage," and has been developed by David Hunt.<sup>35</sup> Here

32. (Mele and Robb's note.) What would happen if Bob's indeterministic deliberative process were to hit N2 at some time  $t_n$  prior to  $t_2$ ? In one version of the story, N2 would light up at  $t_n$  – Bob would decide at  $t_n$  not to steal the car – but then at  $t_2$ , when P hits N1, Bob would change his mind and decide to steal it. In another version – the one we prefer, because of its relative simplicity – P neutralizes N2 as soon as Black initiates P.
33. (Mele and Robb's note.) David Hunt independently makes a similar suggestion in his article, "Moral Responsibility and Unavoidable Action" in *Philosophical Studies* 97 (2000). See note 1.
34. Mele and Robb, "Rescuing Frankfurt-Style Cases," pp. 104–5.
35. Fischer, contribution to a symposium on Kane's *The Significance of Free Will*, *Philosophy and Phenomenological Research* 60 (2000), pp. 141–8; Hunt, "Moral Responsibility and Unavoidable Action."

is a way of presenting this sort of approach that I think is especially powerful. Consider two situations.

**Situation A.** Ms. Scarlet deliberately chooses to kill Colonel Mustard at  $t_1$ , and there are no factors beyond her control that deterministically produce her choice. When she chooses to kill the Colonel, she could have chosen not to kill him. There are no causal factors that would prevent her from not making the choice to kill Colonel Mustard.

In these circumstances, Ms. Scarlet could be morally responsible for her choice. But then, against an alternative-possibilities principle, one might employ a counterfactual version of this situation:

**Situation B.** Ms. Scarlet's choice to kill Colonel Mustard has precisely the same actual causal history as in A. But before she even started to think about killing Colonel Mustard, a neurophysiologist had blocked all the neural pathways not used in Situation A, so that no neural pathway other than the one employed in that situation could be used. Let us suppose that it is causally determined that she remain a living agent, and if she remains a living agent, some neural pathway has to be used. Thus every alternative for Ms. Scarlet is blocked except the one that realizes her choice to kill the Colonel. But the blockage does not affect the actual causal history of Ms. Scarlet's choice, because the blocked pathways would have remained dormant.

One might, at least initially, have the intuition that Ms Scarlet could be morally responsible for her choice in B as well. Yet for an incompatibilist, this intuition might well be undermined on more careful reflection about whether in B Ms Scarlet retains libertarian freedom. One important question about such blockage cases is one Fischer asks: Could neural events bump up against, so to speak, the blockage?<sup>36</sup> If so, there still may be alternative possibilities for the agent. But if not, it might seem, as Kane suggests, that the neural events are causally determined partly by virtue of the blockage.<sup>37</sup>

36. Fischer, "Recent Work on Moral Responsibility," p. 119.

37. Kane suggests this response in his reply to Fischer in the symposium on *The Significance of Free Will in Philosophy and Phenomenological Research* 60 (2000), p. 162: In [a case in which every other alternative is blocked except the agent's making A at  $t$ ], of course, there *are* no alternative possibilities left to the agent; every one is blocked except the agent's choosing A at  $t$ . But now we seem to have determinism pure and simple. By implanting the mechanism in this fashion, a controller would have predetermined exactly what the agent would do (and when); and, as a consequence, the controller, not the agent, would be ultimately responsible for the outcome. Blockage by a controller that rules out all relevant alternative possibilities is simply predestination; and on my view at least, predestination runs afoul of ultimate responsibility.

In response, one might point out that in the standard Frankfurt-style cases, the relevant action is inevitable, but the intuition that the agent is morally responsible for it depends on the fact that it does not have an actual causal history by means of which it is made inevitable. What makes the action inevitable is rather some fact about the situation that is not a feature of its actual causal history, and hence the action's being inevitable need not make it the case that it is causally determined. But then how is the blockage case different from the standard Frankfurt-style cases? After all, the blockage does not seem to affect the actual causal history of the action.

Nevertheless, perhaps Kane's response can be defended. Two-situation cases of the above sort might be misleading just because it is natural to assume that the actual causal history of an event is essentially the same in each, given that the only difference between them is a restriction that would seem to have no actual effect on the event. But consider a simple two-situation case modelled on a reflection of Hunt's.<sup>38</sup> Imagine a universe correctly described by Epicurean physics: At the most fundamental level all that exists is atoms and the frictionless void, and there is a determinate downward direction in which all atoms naturally fall – except if they undergo uncaused swerves.

**Situation C.** A spherical atom is falling downward through space, with a certain velocity and acceleration. Its actual causal history is indeterministic because at any time the atom can be subject to an uncaused swerve. Suppose that the atom can swerve in any direction other than upwards. In actual fact, from  $t_1$  to  $t_2$  it does not swerve.

A counterfactual situation diverges from C only by virtue of a device that eliminates alternative possibilities and all differences thereby entailed:

**Situation D.** The case is identical to C, except that the atom is falling downward through a straight and vertically oriented tube whose interior surface is made of frictionless material, and whose interior is precisely wide enough to accommodate the atom. The atom would not have swerved during this time interval, and the trajectory, velocity, and acceleration of the atom from  $t_1$  to  $t_2$  are precisely what they are in C.

38. From Hunt's personal correspondence with Fischer, cited in Fischer's "Recent Work on Moral Responsibility," pp. 119–20.

One might initially have the intuition that the causal history of the atom from  $t_1$  to  $t_2$  in these two situations is in essence the same. However, this intuition could be challenged by the fact that the restrictions present in D but not in C may change this causal history from one that is essentially indeterministic to one that is essentially deterministic. For since the tube prevents any alternative motion, it would seem that it precludes any indeterminism in the atom's causal history from  $t_1$  to  $t_2$ . And if the tube precludes indeterminism in this causal history, it would appear to make the causal history deterministic. Whether this line of argument is plausible is difficult to ascertain, but it is not obviously implausible.

This problem could make it hard to assess moral responsibility in blockage cases. Sympathy for Frankfurt-style arguments is generated by the sense that moral responsibility is very much a function of the features of the actual causal history of an action, to which restrictions that exist but would seem to play no actual causal role are irrelevant. However, in a scenario in which such restrictions, despite initial appearances, could be relevant to the nature of the actual causal history of an action after all, one's intuitions about whether the agent is morally responsible might become unstable. My own view is not that actual causal histories in blockage cases are clearly deterministic, but only that these considerations show that they may be. This type of problem should make one less confident when evaluating these difficult kinds of Frankfurt-style cases. Since Mele and Robb's development of their case involves something very much like a blockage scenario, one might as a result also be less confident about the ultimate success of their argument.

#### A NEW FRANKFURT-STYLE SCENARIO

I propose a case of a different sort, one that doesn't fit either category (a) or (b):

**Tax Evasion, Part 1.** Joe is considering whether to claim a tax deduction for the substantial local registration fee that he paid when he bought a house. He knows that claiming the deduction is illegal, that he probably won't be caught, and that if he is, he can convincingly plead ignorance. Suppose he has a very powerful but not always overriding desire to advance his self-interest no matter what the cost to others, and no matter whether advancing his self-interest involves illegal activity. Furthermore, he is a libertarian free agent. Crucially, his psychology is such that the only way that in this situation he

could fail to choose to evade taxes is for moral reasons. (As I use the phrase here, “failing to choose to evade taxes” will encompass both not choosing to evade taxes and choosing not to evade taxes.) His psychology is not, for example, such that he could fail to choose to evade taxes for no reason or simply on a whim. In fact, it is causally necessary for his failing to choose to evade taxes in this situation that a moral reason occur to him with a certain force. A moral reason can occur to him with that force either involuntarily or as a result of his voluntary activity (e.g., by his willing to consider it, or by his seeking out a vivid presentation of such a reason). However, a moral reason occurring to him with such force is not causally sufficient for his failing to choose to evade taxes. If a moral reason were to occur to him with that force, Joe could, with his libertarian free will, either choose to act on it or refrain from doing so (without the intervener’s device in place). But to ensure that he chooses to evade taxes, a neuroscientist now implants a device which, were it to sense a moral reason occurring with the specified force, would electronically stimulate his brain so that he would choose to evade taxes. In actual fact, no moral reason occurs to him with such force, and he chooses to evade taxes while the device remains idle.

In this situation, Joe could be morally responsible for choosing to evade taxes despite the fact that he could not have chosen otherwise.<sup>39</sup> The prior sign does not causally determine his decision. There are indeed alternative possibilities which involve a moral reason occurring to him with a certain force. In one type of possibility, Joe makes this happen voluntarily. But such a possibility is insufficiently robust to ground his moral responsibility for tax evasion. Again, the deeper intuition underlying the alternative-possibilities requirement is that if, for example, an agent is to be blameworthy for an action, it is crucial that he could have done something to avoid this blameworthiness. If alternative possibilities were to play a role in explaining an agent’s moral responsibility for an action (in a way independent of an intuition about its actual causal history), it would be because as a result of securing an alternative possibility instead, he would thereby have avoided the responsibility he has for the action he performed. However, if Joe had made a reason for an alternative action occur to him with a certain force, he would not thereby have avoided responsibility for evading taxes. For his making the reason for an alternative action occur to him is compatible with his

39. I pursue this strategy in “Alternative Possibilities and Causal Histories.” Independently, David Hunt, in “Moral Responsibility and Avoidable Action,” pp. 214–16, also considers making the prior sign a necessary condition of the alternative decision, but he is skeptical about this approach. See note 1.



never deciding to perform the alternative action, or even ever being inclined to perform that action, and choosing to evade taxes instead.

It is important to the example that the trigger for intervention be that a moral reason occur to Joe with a certain force, and not simply that a moral reason occur to him. For one might plausibly argue that it is a necessary condition on blameworthiness that the agent understands that his action is morally wrong, which in Joe's case would seem to require some awareness of moral reasons.<sup>40</sup> At the same time, Joe's blameworthiness would not require that moral reasons occur to him with any particularly strong force.

This example fits neither description (a) or (b). Rather, it is a case that has the following features:

- (i) The agent clearly has free will according to most libertarian views.
- (ii) What would trigger the intervention is a "flicker" that is insufficiently robust to explain the agent's moral responsibility for the decision in question.
- (iii) It does not ground the truth of the analog of Widerker's

(1) If Jones is blushing at t1, then, provided no one intervenes, Jones will decide at t2 to kill Smith,

which is

(1') If a moral reason does not occur to Joe with a certain force, then, provided no one intervenes, he will decide to evade taxes,

in causal determinism, while at the same time not endorsing the analog of Widerker's

(1a) If Jones is blushing at t1, then Jones will [only] *probably* decide at t2 to kill Smith,

and

(1b) If Jones is blushing at t1, then Jones will *freely* decide at t2 to kill Smith, [in a sense that allows that the agent could have decided otherwise],

which are

(1a') If a moral reason does not occur to Joe with a certain force, then he will (only) probably decide to evade taxes,

40. Thanks to Michael McKenna for suggesting that I make this point explicit.