

Moral Responsibility, Manipulation Arguments, and History: Assessing the Resilience of Nonhistorical Compatibilism

Michael McKenna

Received: 7 August 2011 / Accepted: 7 August 2011
© Springer Science+Business Media B.V. 2012

Abstract Manipulation arguments for incompatibilism all build upon some example or other in which an agent is covertly manipulated into acquiring a psychic structure on the basis of which she performs an action. The featured agent, it is alleged, is manipulated into satisfying conditions compatibilists would take to be sufficient for acting freely. Such an example used in the context of an argument for incompatibilism is meant to elicit the intuition that, due to the pervasiveness of the manipulation, the agent does not act freely and is not morally responsible for what she does. It is then claimed that any agent's coming to be in the same psychic state through a deterministic process is no different in any relevant respect from the pertinent manner of manipulation. Hence, it is concluded that compatibilists' proposed sufficient conditions for free will and moral responsibility are inadequate, and that free will and moral responsibility are incompatible with determinism. One way for compatibilists to resist certain manipulation arguments is by appealing to historical requirements that, they contend, relevant manipulated agents lack. While a growing number of compatibilists advance an historical thesis, in this paper, I redouble my efforts to show, in defense of nonhistorical compatibilists like Harry Frankfurt, that there is still life left in a nonhistorical view. The historical compatibilists, I contend, have fallen shy of discrediting their nonhistorical compatibilist rivals.

I presented earlier drafts of this paper to the philosophy departments at University of Arizona and University of Tennessee. I received much excellent feedback from members of both audiences. For helpful advice, I am indebted to Mark Balaguer, Randy Clarke, E. J. Coffman, Ish Haji, Terry Horgan, Stephen Kearns, Keith Lehrer, Al Mele, Shaun Nichols, David Owen, David Palmer, Derk Pereboom, Guido Pincione, Carolina Sartorio, David Schmidtz, Jim Stone, Steve Wall, and Brandon Warmke. Finally, I would like to thank Angelo Corlett for inviting me to contribute to this special issue on moral responsibility.

M. McKenna (✉)
Department of Philosophy, Center for the Philosophy of Freedom,
University of Arizona, Tucson, AZ, USA
e-mail: msmckenna63@gmail.com

Keywords Derivative freedom · Direct freedom · Harry Frankfurt · Historical compatibilism · Incompatibilism · Nonhistorical compatibilism · Manipulation argument · Alfred Mele

Manipulation arguments for incompatibilism all build upon some example or other in which an agent is covertly manipulated into acquiring a psychic structure on the basis of which she performs an action. The featured agent, it is alleged, is manipulated into satisfying conditions compatibilists would take to be sufficient for acting freely. Such an example used in the context of an argument for incompatibilism is meant to elicit the intuition that, due to the pervasiveness of the manipulation, the agent does not act freely and is not morally responsible for what she does. It is then claimed that any agent's coming to be in the same psychic state through a deterministic process is no different in any relevant respect from the pertinent manner of manipulation. Hence, it is concluded that compatibilists' proposed sufficient conditions for free will and moral responsibility are inadequate, and that free will and moral responsibility are incompatible with determinism.

Although manipulation arguments have been in circulation for some time, recently they have taken center stage in current debates surrounding the topics of free will and moral responsibility.¹ I have contributed to these debates with two articles that have come in for a good deal of thoughtful critical scrutiny.² Each article was devoted to a different task. In one, I set out what in my estimation is the best overall strategy for a compatibilist when responding to a manipulation argument.³ There, as I shall explain below, I advised compatibilists to favor what I termed a hard-line reply over a soft-line reply wherever possible. In a distinct paper, I took up what is perhaps largely, albeit not exclusively, an in-house controversy amongst compatibilists as to whether moral responsibility is essentially an historical concept or instead a nonhistorical one (also to be explained below).⁴ While a growing number of compatibilists advance an historical thesis, I argued that there is still life left in a nonhistorical view.⁵ The historical compatibilists, I contended, had fallen shy of discrediting their nonhistorical compatibilist rivals. The dispute is a delicate one, since an historical thesis appears to give compatibilists extra resources for responding to certain formulations of manipulation arguments.

¹ Earlier formulations include Taylor (1974), and Locke (1975). More recently, see Kane (1996), Pereboom (2001), and Mele (2006).

² See Demetriou (2010), Haji (2009), Haji and Cuypers (2004, 2007), Mele (2008, 2009b), and Pereboom (2008).

³ McKenna (2008a).

⁴ McKenna (2004).

⁵ Historical compatibilists include Fischer and Ravizza (1998), and Haji (1998). Mele advises compatibilists to adopt an historical view, but remains agnostic about compatibilism; see Mele (1995, 2006). Nonhistorical compatibilists include Arpaly (2006), Berofsky (2006), Double (1991), Dworkin (1988), Frankfurt (1975), Vargas (2006), Watson (1999), and Wolf (1987).

In this paper, I focus primarily on the debate between historical and nonhistorical compatibilists. My goal is to redouble my earlier efforts to show that nonhistorical compatibilism remains a live option. In setting out the nonhistorical compatibilists' case, I hope to bring a greater degree of clarity to the debate. For this reason, I shall proceed slowly, attending initially to the structure that manipulation arguments take, and then to the content of what an interesting historical thesis should come to. With all of this in place, I will explore the resources available to the nonhistorical compatibilist.

1 Dialectical Preliminaries

I begin with some stage-setting. *Incompatibilism*, as I use the term, is the thesis that both free will and moral responsibility are incompatible with determinism; *compatibilism* is the thesis that both are compatible with determinism. Although controversial, I define *free will* as the unique ability of persons to control their conduct in the strongest manner necessary for moral responsibility. A *free act* just is one that issues from a non-deviant exercising of that ability. For present purposes, it will suffice to characterize *moral responsibility* in terms of praiseworthiness and blameworthiness for actions. Thus, we can make do with the following proposal: an agent is *blameworthy* for an act *x* if she freely performs *x* knowing that *x* is morally wrong, and an agent is *praiseworthy* for an act *x* if she freely performs *x* knowing that *x* exceeds what is morally required.⁶ For argument's sake, I will assume along with Derk Pereboom⁷ a *basic-desert entailing* notion of moral responsibility such that an agent deserves blame just because she performed a blameworthy act, or instead deserves praise just because she performed a praiseworthy act, and not because of any further considerations such as those a consequentialist or a contractualist might enlist.⁸ Finally, *determinism* can be characterized as “the thesis at any instant there is exactly one physically possible future.”⁹

Strictly speaking, given the preceding definitions, the debate between compatibilists and incompatibilists is simply a debate about what is metaphysically possible. Compatibilism is vindicated if it can be established that there exists at least one deterministic world at which at least one agent performs one free act for which she is morally responsible. But it would be a disappointing compatibilist victory if the compatibilist were not able to establish more than that. A successful compatibilism, I shall assume, must also make credible that actual human persons—in particular, most psychologically healthy, normally functioning adults—do act freely and are morally responsible for a good deal of their morally significant conduct. In assessing manipulation arguments for incompatibilism, I will therefore assume that an adequate

⁶ My view (2011) is that these sufficient conditions are inadequate, but that need not detain us here.

⁷ Pereboom (2001, p. xx).

⁸ The blame or praise that is at issue, as I understand it, involves some overt bit of behavior or treatment, and not merely a private attitude of blame or praise.

⁹ van Inwagen (1983, p. 3).

compatibilist reply must make do with only plausible assumptions about human psychology.

One noteworthy assumption about human psychology which will figure prominently in the ensuing discussion has to do with an agent's relation to her own values. Here I shall follow Alfred Mele in thinking of valuing, or as he would put it, *thinly valuing*, as follows: "S at least *thinly values* X at a time if and only if at that time S both has a positive motivational attitude toward X and believes X to be good."¹⁰ So construed, valuing is quite inclusive—hence Mele's modifier 'thin'. I have a positive motivational attitude toward peanut butter, and I believe it to be good (e.g., healthy). Ergo, I value peanut butter. Nevertheless, this definition will also bring within its sweep the more portentous kind of valuing that often figures prominently in assessing a person's moral praiseworthiness and blameworthiness. Sam might value torturing and the degradation of certain minorities, and Sally might value liberal education and the welfare of her children. Now it might well be that on some occasions a person acts freely and is morally responsible for some act in which none of her values play any role at all. But I think it is nearly beyond dispute that by and large most actual psychologically healthy adult persons on many occasions, and most especially occasions that are morally charged, act from some values that are of the more significant sort—that is, are not merely peanut-butter valuing but are welfare-of-ones-children valuing. Of course, sometimes persons act contrary to something they value. They act weakly, for instance. But often persons act from, and not merely in accord with, their values. That is to say, their values play some motivational—and so, on my view, causal—role in their acting as they do.¹¹

I suppose some of the values a person possesses at a time are ones that it is up to her whether she possesses them at that time; were it so, she would be able to lose or substantially downgrade them at will. But I strongly suspect that most values that most actual persons possess at a time are such that, at that very time, those values are *practically unsheddable* for them.¹² While they might be able to set themselves on a course to rid themselves of those values at some point in the future, perhaps through therapy or some process of education or indoctrination, at that time their values are ones which are such that, in normal practical contexts, it is simply not up to them whether they possess those values, nor is it up to them the relative strength those values have for them. For instance, it is not up to me right now that I value my wife's welfare or my own health as much as I do. Save for very exotic conditions outside the sphere of normal contexts of deliberation and action, they are values that I cannot but possess with the strength that they have for me. In assessing manipulation arguments, I will therefore restrict my attention to arguments in which

¹⁰ Mele (1995, p. 116).

¹¹ I will operate with this notion of value, though it does not fully capture the relevant class of attitudes. One could desire X and think it good and yet not be motivated to X *because* one was committed to its value (on a related point, see Watson 2002: 144–145). Judging X valuable, even when motivated to pursue X, is more inclusive than valuing X. When speaking of a person's values as they figure in explanations of her actions, presumably we are interested in the more restrictive notion. I will set this aside here, as it will not bear on the discussion to follow.

¹² The useful notion of unsheddable values is introduced in Mele (1995, p. 153; and 2006, pp. 167–168).

the manipulated agents featured in them come equipped not just with values, but with unsheddable values.

2 The Manipulation Argument

I started by writing in terms of manipulation *arguments*, which is meant to indicate that there are varying instances of them. But it is also useful to think of *the* manipulation argument, where this refers not to any specific argument, but rather to an argument form. Instances are then generated from the particular examples that figure in them. The examples involve cases in which agents are, as I have termed it elsewhere, *globally* rather than just *locally* manipulated.¹³ That is, a featured agent is massively tinkered with in such a way as to alter or fix substantially a large swath of her psychological states or characteristics; she is not merely caused to acquire, say, just a particular belief or desire.¹⁴ So, to proceed, treat 'X' as a placeholder for whatever manner of global manipulation is to be substituted for it. A substitution instance for manner X is usually meant by an incompatibilist to satisfy some targeted set of proposed conditions at least minimally sufficient for what might be called the Compatibilist-friendly Agential Structure (CAS). When satisfied and acted upon, CAS is supposed to be minimally sufficient for acting freely and being morally responsible even in conditions in which determinism is true. The particular content of CAS featured in an incompatibilist's manipulation argument will depend upon which compatibilist proposal an incompatibilist has in her sites. It might be A. J. Ayer's, Harry Frankfurt's, John Martin Fischer and Mark Ravizza's, or R. Jay Wallace's.¹⁵ Alternatively, it might be the conjunction of all four, which is how Pereboom formulated his version of a manipulation argument.¹⁶ Given some example or other so structured, the argument form can be set out as follows:

1. If an agent, S, is manipulated in manner X to perform act A, then S does not A freely and is therefore not morally responsible for A-ing.
2. Any agent manipulated in manner X to A is no different in any relevant respect from any normally functioning agent determined to do A from CAS.
3. Therefore, any normally functioning agent determined to do A from CAS does not A freely and therefore is not morally responsible for A-ing.

The first premise, properly filled out, is supposed to be established on the basis of intuition. The second, as I understand it, is meant as an assertion that should be accepted as true in the absence of any credible reason as to why a mere determined history is different in a relevant way from some unusual causal process brought

¹³ McKenna (2004, p. 169).

¹⁴ A background assumption is that the manipulation is not so excessive as to violate conditions of the agent's identity over time. Hence, on certain psychological continuity views of personal identity, a coherence constraint on the manipulation story will require a certain level of psychological connectedness between pre- and post-manipulation. In what follows, I will simply assume that these conditions, whatever they come to, can be satisfied.

¹⁵ See Ayer (1954), Frankfurt (1971), Fischer and Ravizza (1998), and Wallace's (1994).

¹⁶ See Pereboom (2001, p. 113).

about through manipulation. The conclusion follows, perhaps with the aid of a further enthymematic premise that like cases should be treated alike.^{17,18}

To illustrate, consider Robert Kane's characteristically thoughtful formulation of a particular instance of the manipulation argument by use of what he calls *Covert Nonconstraining Control* (CNC control). Kane writes:

We are well aware of these two ways to get others to do our bidding in everyday life. We may force them to do what we want by coercing or constraining them against their will, which is constraining or CC control. Or we may manipulate them into doing what we want while making them feel that they have made up their own minds and are acting “of their own free wills”—which is covert nonconstraining or CNC control. Cases of CNC control in larger settings are provided by examples of behavioral engineering such as we find in utopian works like Aldous Huxley's *Brave New World* or B. F. Skinner's *Walden Two*. Frazier, the fictional founder of Skinner's Walden Two, gives a clear description of CNC control when he says that in his community persons can do whatever they want or choose, but they have been conditioned since childhood to want and choose only what they can have or do.

Utopian versions of CNC control such as B. F. Skinner's are an excellent medium for expressing [an] objection to hierarchical views of free will such as Frankfurt's. For the citizens of Walden Two, as they are described, are marvelously “wholehearted” in their attitudes and engagements. They are “satisfied” with themselves (almost to a cloying degree) and “have the wills they want to have.” As a result, Frazier, the founder of Walden Two, insists that it is “the freest place on earth.” And why should it not be? There is no coercion in Walden Two and no punishment. All citizens can have and do whatever they want or choose; and they can *will* to do whatever they *want*—that is, their

¹⁷ This formulation is largely the same as my earlier efforts, save for one important amendment. Here I write more cautiously of CAS satisfying some proposed compatibilist minimally sufficient conditions, rather than of it “exhausting” compatibilist conditions (McKenna 2004, p. 169; and 2008a, p. 143). Mele objected to my earlier formulation (Mele 2009b, pp. 467–469), where I wrote of CAS satisfying “the very richest” of conditions for freedom and responsibility (McKenna 2004, p. 182). That was too excessive, though I was merely alluding to the relatively robust conditions Pereboom operates with, the conjunctive one mentioned above.

¹⁸ This is the only argument form I will consider. But note that Mele (2008) distinguishes three different forms. The first, which he calls a *straight manipulation argument*, is basically the form I have set out. The second form, a *manipulation argument to the best explanation*, which is due to Pereboom (2001), replaces the second premise as I have set it out above with a best-explanation premise of roughly the form, ‘the best explanation that S is not free and morally responsible when manipulated in manner X into A-ing is that S is causally determined.’ The third form, an *original design argument* of the sort Mele himself has developed (2006), works not from a case of an already existing agent who is manipulated, but from an agent who is originally designed to perform some act, A. Mele is correct to mark these distinctions. However, it is not clear to me how much the distinction between the first and third comes to. Originally designing an agent to do something at a later time seems also to be a case of manipulation, but a case in which the agent is manipulated from a point in time prior to her coming into existence. Kearns (2011) makes a similar point.

first-order desires always conform to their second-order volitions. They have maximal freedom of will and action in the hierarchical sense.

Yet most people who look at Walden Two would say—and say rightly, in my opinion—that its citizens lack free will in a deeper sense than being able to do what they want and will what they want. In this deeper sense, their wills are not “their own” because they are not the original creators of their own ends or purposes. Their ends or purposes are created by their conditioners or controllers.¹⁹

The preceding way of presenting a manipulation argument, as is illustrated by Kane’s formulation, is limited to discrediting existing compatibilist proposals—in so far as CAS is tailored to some specific set of proposed sufficient conditions. Kane’s target was, obviously, Frankfurt’s proposal. But as Mele points out, an incompatibilist might have more ambitious aims meant to show not merely that some compatibilist proposal is false, but rather that incompatibilism is true, full stop.²⁰ Hence, a bolder version of the manipulation argument might be formulated as follows²¹:

1. If an agent, S, is manipulated in manner X to perform act A, then S does not A freely and is therefore not morally responsible for A-ing.
2. Concerning free action and moral responsibility, there is no significant difference between S’s A-ing as a result of being manipulated into doing so in manner X and any candidate for a free and morally responsible action in a deterministic universe.
3. Therefore, any normally functioning agent determined to do A does not A freely and therefore is not morally responsible for A-ing.

Throughout, I will assume that particular manipulation arguments up for assessment will include bolder formulations of this sort.

As I have noted elsewhere, there is no one-size-fits-all compatibilist reply to manipulation arguments.²² Which premise a compatibilist should reject in a particular instance of a manipulation argument will depend upon that compatibilist’s considered views in relation to the example featured in the argument. A compatibilist who works from a sufficiently nuanced theory of freedom might well be able to detect some way that a particular case of manner-X manipulation fails to satisfy her preferred formulation of CAS. She would have the resources to reject the second premise while leaving the first well enough alone. On the other hand, if instead the manner of manipulation is crafted so as truly to nail down a compatibilist’s proposed sufficient conditions, that compatibilist is under pressure to accept the second premise of the pertinent version of a manipulation argument and instead reject the first.²³ Hence, I

¹⁹ Kane (1996, p. 65).

²⁰ Mele (2008, pp. 264–265).

²¹ This formulation draws upon Mele’s, especially the second premise (2008, p. 265).

²² McKenna (2008a, p. 143).

²³ It is also open to her to amend her version of CAS further in some fashion that is not blatantly *ad hoc*, thereby allowing her to resist the second premise. But of course, if she makes this move, the incompatibilist is certainly entitled to parry with a counter-move of revising the case of manipulation so as to suit it for the compatibilist’s amended formulation of CAS.

distinguished between *hard-line replies* and *soft-line replies* to instances of the manipulation argument.²⁴ A hard-line reply rejects the first premise of a manipulation argument; a soft-line reply rejects the second. What makes one reply hard and another soft is that in the former case, the compatibilist has to take on a hard task. She must argue against an assertion that seems to have a considerable amount of intuitive support behind it; on its face, an agent massively manipulated into conditions ensuring a particular course of action seems not to act freely. In short, a hard-line reply commits a compatibilist to bullet-biting in a way that a soft-line reply does not. Appreciating this point helps to assess what is at stake in the dispute between historical and nonhistorical compatibilists, as I shall now explain.

3 Manipulation Arguments and History

To help bring into relief what is at issue in the historical/nonhistorical debate, consider the following two passages from the nonhistorical compatibilist Harry Frankfurt:

To the extent that a person identifies himself with the springs of his actions, he takes responsibility for those actions and acquires moral responsibility for them; moreover, the questions of how the actions and his identifications with their springs are caused are irrelevant to the questions of whether he performs his actions freely and is morally responsible for performing them.

And:

What we need most essentially to look at is, rather, certain aspects of the psychic structure that is coincident with the person's behavior.²⁵

Briefly, it seems to me that if someone does something because he wants to do it, and if he has no reservations about that desire but is wholeheartedly behind it, then—so far as his moral responsibility is concerned—it really does not matter how he got to be that way. One further amendment must be added to this: the person's desires and attitudes have to be relatively well integrated into his general psychic condition. Otherwise they are not genuinely his, but are merely disruptive intruders on his true nature. As long as their interrelations imply that they are unequivocally attributable to him as his desires and attitudes, it makes no difference—so far as evaluating his moral responsibility is concerned—how he came to have them.

²⁴ McKenna (2008a, p. 143). This distinction is similar to Kane's distinction between hard compatibilism and soft compatibilism (Kane 1996, pp. 67–68), which is also adopted by Watson (1999, pp. 360–365). I fear that their way of marking the distinction has mistakenly been taken to suggest that compatibilists must adopt a one-size-fits-all stance towards manipulation arguments. I should add that I do not believe either of them made that mistake. In any event, the terminology I have proposed is meant to guard against that suggestion.

²⁵ Frankfurt (1975, as appearing in 1988, p. 54).

A manipulator may succeed, through his interventions, in providing a person not merely with particular feelings and thoughts but with a new character. That person is then morally responsible for the choices and the conduct to which having this character leads. We are inevitably fashioned and sustained, after all, by circumstances over which we have no control. The causes to which we are subject may also change us radically, without thereby bringing it about that we are not morally responsible agents. It is irrelevant whether those causes are operating by virtue of the natural forces that shape our environment or whether they operate through the deliberate manipulative designs of other human agents. We are the sorts of persons we are; and it is what we are, rather than the history of our development, that counts. The fact that someone is a pig warrants treating him like a pig, unless there is reason to believe that in some important way he is a pig against his will and is not acting as he would really prefer.²⁶

In the first passage quoted, Frankfurt is responding to Don Locke's challenge put to Frankfurt that a Devil/neurologist could manipulate an agent to be in a state that would satisfy Frankfurt's proposed conditions for CAS.²⁷ In the second, he is responding to the historical compatibilist Fischer's formulation of the same point.²⁸ As one can plainly see, Frankfurt's defiant response would also be perfectly suited as a response to the quotation from Kane enlisted above.

Historical compatibilists, in clear opposition to Frankfurt's contention, have argued that among the conditions necessary for free and morally responsible action is some kind of history requirement. Either the agent must have come to be in the state she is in just prior to her free action through some kind of freedom-and-responsibility-enabling historical process,²⁹ or she must *not* have come to be in the pertinent state from a history that included any freedom-and-responsibility-undermining process.³⁰ Call the former sort *positive historical compatibilists* and the latter sort *negative historical compatibilists*. For instance, Fischer and Ravizza advance a positive historical thesis.³¹ They require, roughly, that when an agent acts freely, she acts from agential springs that are hers—that is, that she *owns*—by virtue of the fact at an earlier time she took responsibility for those springs of action. She did this by going through a process whereby she came to see herself as an agent in the world, one who is a fair target of blame and praise, and she came to so see herself as such on the basis of appropriate evidence. Agents who have not come to acquire their springs of action at a time through such a history do not act freely while, other things being equal, agents who have such a history do act freely. Mele instead recommends a negative historical thesis.³² Unlike the positive historical

²⁶ Frankfurt (2002, pp. 27–28).

²⁷ Locke (1975).

²⁸ Fischer (2002).

²⁹ See, for example, Fischer and Ravizza (1998), or Haji (1998).

³⁰ For example, see Mele (1995, 2006).

³¹ Fischer and Ravizza (1998, pp. 207–214).

³² Mele (1995, Ch. 9, pp. 144–176; 2006, pp. 164–173).

compatibilists, on the view he advises compatibilists to adopt, there is no particular kind of history an agent must possess relative to her actions for them to be free. An agent acts freely and is morally responsible for what she does only if she *lacks* a history of a certain sort relative to the causes of her action. Roughly, if her act is generated in a certain way by pertinent attitudes, such as unsheddable values, and if she came to acquire those attitudes through coercive means, means that bypassed her ability to critically assess, endorse, and sustain them, she does not act freely.³³

Historical compatibilists appear to be at an advantage in comparison with their nonhistorical adversaries. In many manipulation arguments, such as Kane's presented above, the manipulation cases featured in them will be such that the agents are manipulated so that they do not have the sorts of histories that the positive historical compatibilists would require for freedom and responsibility. Or, instead, they are manipulated such that they *do* have the sorts of histories that the negative historical compatibilists would take to be freedom and responsibility defeating. Yet these manipulated agents might very well satisfy all that a nonhistorical compatibilist, such as Frankfurt, demands for his or her account of CAS. In this wide range of cases, unlike the nonhistorical compatibilists, the historical compatibilists will (putatively) have the resources to resist the respective second rather than the respective first premises of this range of arguments. That is, for any instance in this range, the historical compatibilists will be in a position to grant that the manipulated agent is not free and responsible. Instead, the historical compatibilist will be able to take a soft-line reply and contend that there *is* a relevant difference between such manipulation and the way mere determinism causes free and responsible agents to act. Some histories, they will be able to argue, are freedom and responsibility-defeating; others are not. This would count as an advantage for the historical compatibilists because, as noted in the preceding section, taking a hard-line reply and denying the first premise of a manipulation argument comes with the risk of bullet-biting. There is, at least in many of the cases in circulation, considerable intuitive pull in the direction of judging that the featured manipulated agents are not free and responsible for what they do.

³³ As many readers will no doubt recognize, the dispute between nonhistorical and historical compatibilists is yet another variation on the internalism/externalism debate that crops up in all kinds of places. Nonhistorical compatibilists like Frankfurt, as illustrated in the passages quoted above, are committed to the view that any two agents who are nonhistorical duplicates by virtue of sharing all of their "nonhistorical," "snapshot" or "current timeslice properties" do not differ with respect to their status as free and morally responsible (as I shall explain momentarily, this has to be qualified in terms of direct freedom and direct responsibility). Historical compatibilists deny this; two agents who are nonhistorical duplicates at a time might very well differ with respect to their status as free and morally responsible depending upon differences in their respective histories—that is, depending upon differences in their "historical properties". Hence, for the historical compatibilists, the concept of moral responsibility is historical in the same way that the property of being a sunburn or a genuine dollar bill is historical. Two burns at a time might be qualitatively indistinguishable. But because only one was caused by the sun and another not, only one is a sunburn. The same with two distinct one dollar bills; if one is counterfeit and the other produced properly, only the latter is genuine. The nonhistorical compatibilist, by contrast, is committed to the view that free will and moral responsibility are nonhistorical in the same way that the property of being a certain size, shape or weight is nonhistorical; any two duplicates sharing the same size, shape, or weight at a time are qualitatively identical in the pertinent respect regardless of any differences in how they came to be that way. See Fischer and Ravizza for a clear articulation of these notions (1998, pp. 171–173).

Thus, one way to appreciate the debate between the nonhistorical compatibilists and their historical adversaries is with respect to the resources compatibilists have available to them for responding to a certain range of manipulation arguments. Seen in these terms, the debate boils down to whether the historical compatibilists are entitled to help themselves to the presumption of the historical constraints that they endorse. If they are, then they make the dialectical burden for compatibilism lighter with respect to a considerable range of manipulation arguments. Hence, is free and morally responsible action really historical in the way that the historical compatibilists claim that it is? In attempting to assess this debate, I have argued that it is not clear that the nonhistorical compatibilist position has been put to rest.³⁴ While I have not committed to the nonhistorical thesis and have officially remained agnostic as to where I stand, I have argued that the nonhistorical view is less vulnerable than some historical compatibilists take it to be.³⁵ Before turning directly to my assessment of this controversy, I pause to get a bit clearer on just what a credible and philosophically interesting historical thesis comes to.

4 Zeroing in on a History Thesis

An especially important qualification in the dispute between historical and nonhistorical compatibilists, one that seems not to have been fully integrated into the dialectic, turns on a distinction between *direct freedom* and *derivative freedom*.³⁶ Above, I characterized a free act as one that issues from an ability which is unique to persons whereby an agent controls her conduct in the strongest manner necessary for moral responsibility. But free acts can be more finely specified. Some gain their status as free directly from an exercise of the free will ability which issues in those very acts; others are such that their status as free is entirely derivative and so depends on standing in the proper kind of causal relations to prior acts which themselves are directly free acts. A parallel distinction is also drawn for moral responsibility. There are cases of *direct moral responsibility* and *derivative moral responsibility*. While I do not think it is fully accurate to think that the latter distinction is to be accounted for exclusively in terms of the former, in the present context, there is little harm in operating as such.³⁷ Thus, I will assume that

³⁴ McKenna (2004).

³⁵ McKenna (2004, p. 190).

³⁶ Alternative terminology sometimes used is *direct* and *indirect*, or instead, *basic* and *non-basic*.

³⁷ The main reason, although not the only reason, I do not think it is fully accurate is because it seems to me that some cases are better described as cases where a person is not even derivatively free but is derivatively morally responsible, and her derivative moral responsibility traces back to prior directly free acts for which she was directly morally responsible. Such cases might be ones such as reckless drug use leading to unwanted addiction. Cases where the derivative responsibility coincides with derivative freedom are probably best limited to cases where at the time of the behavior in question, the agent welcomes being in the state she is in, despite the fact that, while in it, it is no longer up to her how she will act or what will result from her prior directly free undertakings. Tending to this variation would require various amendments and modifications to the discussion throughout that would not affect the main line of argumentation. So for ease of discussion, I will just treat the notions of derivative freedom and derivative moral responsibility as a package deal.

when an agent's moral responsibility for an action is derivative rather than direct, this will be due to the fact that the act is only derivatively free and that its moral responsibility status tracks that.

Transparent cases of derivative freedom and moral responsibility are drunk-driving cases and Ulysses cases. An agent might freely embark upon the plan of both getting drunk and then driving home in that state. Here it is reasonable to think that she drives home freely and is morally responsible for doing so, and this can be true even if it is also true that at the time she drove home, she had no idea what she was doing and was then unable to exercise any direct control over her conduct. It is the distinction between direct and derivative freedom and responsibility that warrants such a judgment. Then there is the case of Ulysses, who freely chose to tie himself to the mast so that soon thereafter he could hear the sirens sing. In this case, when later the sirens sang, Ulysses freely listened, and was responsible for placing himself in the state he was in. But at the time he was not directly free. Indeed, he was coerced—by his former self.

The point I now wish to make is that I do not think it is even remotely credible for any compatibilist—or for that matter, any theorist about free will or moral responsibility—to advance a theory that does not make room for the distinction between direct and derivative freedom and the related distinction between direct and derivative moral responsibility. But notice that the distinction just is an historical one. Hence, I assert, *all* compatibilists, nonhistorical compatibilists included, should be committed to *some* historical presuppositions about free will and moral responsibility. And, furthermore, this is a completely trivial and non-controversial historical thesis. Now perhaps it is true that someone like Frankfurt does not make any such distinction at all—if so, so much the worse for Frankfurt's view.³⁸ The interesting philosophical controversy up for dispute, I contend, has to do with whether directly free acts, ones for which an agent is directly morally responsible, are such that a history condition is required *for them*.³⁹ It is just obvious that two agents qualitatively identical in being roaring drunk are such that only one is

³⁸ In my estimation, Frankfurt can readily take on the distinction while preserving his view. It is true that some of his remarks suggest that he would not make room for indirect freedom and responsibility, in particular, his discussion of the unwilling addict, which he treats as transparently not free and not responsible (1971). This is precisely the kind of case where it would be natural to invoke the distinction for some cases of unwilling addiction and not others. It just seems incredible that Frankfurt would not agree that a person who at an earlier time freely engaged in drug use knowing the moral hazards she risked could not later be morally responsible for her addictive behavior even if, once she became an addict, she was an unwilling one. Frankfurt's own considered position aside, it is easy enough to see how one could fashion a "Frankfurtian" view that includes the distinction between direct and derivative freedom and responsibility.

³⁹ The distinction I am drawing upon here is not touched by Mele's complaint about a different way of formulating the notion of basically free actions. According to the formulation that Mele considers and rejects, basicness is defined simply in terms of an act whose status as free is history-independent (2009b, p. 469). All such actions, Mele points out, are by definition history-independent, and so would yield only trivial results for one seeking to defend a nonhistorical view. Indeed. But the distinction I make turns on actions whose status as free is independent of a *particular kind of history*—in particular, a history that includes prior free acts by virtue of which later ones gain their status as free. It is an open question, and so not trivial, for *this* class of acts—the directly free ones—as to whether their status as free depends upon any *further* historical requirements.

responsible and not the other if only one freely chose to get drunk and then drive while the other was force-fed two bottles of tequila and put behind the wheel of a car. Only the former is morally responsible for any ensuing disaster.

Thus, the historical/nonhistorical debate is—or at any rate, in my opinion, *ought* to be—about directly free actions for which an agent is directly morally responsible. Understood in these terms, it is useful to think about the dispute as follows:

Compatibilist theories are meant at the very least to offer up conditions minimally sufficient for *directly* free action. Some compatibilists propose that free will be understood in terms of the ability to do otherwise;⁴⁰ others instead propose that it be understood in terms of responsiveness to reason;⁴¹ yet others appeal to structural features of an agent's psychic architecture.⁴² There are other options. Typically, whatever the theory is on offer, the compatibilist advancing it attempts some more precise analysis or elucidation of the basic proposed ingredients. So, for instance, Ayer analyzed the ability to do otherwise in terms of simple counterfactuals,⁴³ and Frankfurt analyzed an agent's willing what she wants to will in terms of effective first-order desires, a class of higher-order desires (volitional desires), and the notion of identification.⁴⁴ Whatever the conjunctive constellation of conditions a compatibilist sets out to account for her preferred theory *as applied to directly free acts*, historical compatibilists argue that amongst them, one must be an historical condition, either a positive historical condition or a negative one. Nonhistorical compatibilists deny this.⁴⁵

So as to proceed economically, in considering cases meant to test the controversy between historical and nonhistorical compatibilists, I will assume that the manipulated agents featured in them could have done otherwise. And I will also assume that 'could have done otherwise' is itself a nonhistorical notion. Here I intend 'could have done otherwise' to do double duty by functioning disjunctively. Either it refers to the ability to do otherwise as a compatibilist might understand it, or it functions as a stand-in for other credible compatibilist proposals, such as a view in which an agent who acts freely and is morally responsible is somehow suitably reasons-responsive despite not being able to do otherwise (the latter is the sort of view that I favor). Let it be granted that in these cases, because this

⁴⁰ For example, see Ayer (1954), Berofsky (1987), Campbell (1997), Horgan (1985), Lehrer (1976), Lewis (1981), Smith (2003) or Vihvelin (2004).

⁴¹ For example, see Fischer and Ravizza (1998), Gert and Duggan (1979) or Haji (1998).

⁴² For example, see Dworkin (1988), Frankfurt (1971), or Watson (1975).

⁴³ Ayer (1954).

⁴⁴ Frankfurt (1971).

⁴⁵ A variation on this strategy is one in which it is simply granted that the freedom condition of morally responsible action is itself nonhistorical, but then some further condition, say an authenticity or autonomy condition—is required for direct moral responsibility, and it is this condition that involves some historical consideration. This is how Haji argues for the positive historical thesis that he has developed (1998). Purely as a way of simplifying matters of bookkeeping, I will assume that the freedom condition can be treated in an inclusive manner so as to also involve authenticity or autonomy as someone like Haji might think of it.

could-have-done-otherwise condition is met, the manipulated agents satisfy whatever nonhistorical properties both sides can agree figure in a respectable compatibilist theory of directly free actions.

The preceding formulation is especially important in light of my earlier discussion of Mele.⁴⁶ There I focused upon Mele's much discussed case involving Ann and Beth.⁴⁷ Here is Mele's more recent recounting of the case, which is the one I commented upon:

Ann is a free agent and an exceptionally industrious philosopher. She puts in twelve solid hours a day, seven days a week, and she enjoys almost every minute of it. Beth, an equally talented colleague, values many things above philosophy for reasons that she has refined and endorsed on the basis of careful critical reflection over many years. Beth identifies with and enjoys her own way of life, and she is confident that it has a breadth, depth, and richness that long days in the office would destroy. Their dean wants Beth to be like Ann. Normal modes of persuasion having failed, he decides to circumvent Beth's agency. Without the knowledge of either philosopher, he hires a team of psychologists to determine what makes Ann tick and a team of new-wave brainwashers to make Beth like Ann. The psychologists decide that Ann's peculiar hierarchy of values accounts for her productivity, and the brainwashers instill the same hierarchy in Beth while eradicating all competing values—via new-wave brainwashing, of course. Beth is now, in the relevant respect, a “psychological twin” of Ann. She is an industrious philosopher who thoroughly enjoys and highly values her philosophical work. Largely as a result of Beth's new hierarchy of values, whatever upshot Ann's critical reflection about her own values and priorities would have, the same is true of critical reflection by Beth. Her critical reflection, like Ann's, fully supports her new style of life.

Naturally, Beth is surprised by the change in her. What, she wonders, accounts for her remarkable zest for philosophy? Why is her philosophical work now so much more enjoyable? Why are her social activities now so much less satisfying and rewarding than her work? Beth's hypothesis is that she simply has grown tired of her previous mode of life, that her life had become stale without her recognizing it, and that she finally has come fully to appreciate the value of philosophical work. When she carefully reflects on her values, Beth finds that they fully support a life dedicated to philosophical work, and she wholeheartedly embraces such a life and the collection of values that supports it.

Ann, by hypothesis, freely does her philosophical work; but what about Beth? In important respects, she is a clone of Ann—and by design, not an accident. Her own considered values were erased and replaced in the brainwashing process. Beth did not consent to the process. Nor was she even aware of it; she had no opportunity to resist. By instilling new values in Beth and eliminating

⁴⁶ McKenna (2004).

⁴⁷ Mele (1995, pp. 145–146).

old ones, the brainwashers gave her life a new direction, one that clashes with the considered principles and values she had before she was manipulated. Beth's autonomy was violated. And it is difficult not to see her now, in light of all this, as heteronomous—and unfree—to a significant extent in an important sphere of her life. If that perception is correct, then given the psychological similarities between the two agents, the difference in their current status regarding freedom would seem to lie in how they *came* to have certain of their psychological features, hence in something *external* to their present psychological constitutions. That is, the crucial difference is *historical*; free agency is in some way history-bound.⁴⁸

I find the case of Ann and Beth to be an especially useful one for thinking about the historical/nonhistorical debate. Unfortunately, in my earlier treatment, I misrepresented Mele's considered view about how his Ann and Beth case is to be understood.⁴⁹ I had assumed from Mele's telling of the above story that after the manipulation Beth had come to acquire a collection of unsheddable values upon which, like Ann, she subsequently acted. I was right in thinking this.⁵⁰ But I had also assumed, wrongly according to Mele,⁵¹ that Beth could have done otherwise—in a respectable compatibilist sense of that notion—with respect to doing her philosophical work and with respect to the new direction that her life took.⁵²

Unfortunately, if we are to interpret the case of Ann and Beth the way Mele thinks of it, I do not think it will be useful to draw upon it in the service of settling what I take to be the most interesting debate between nonhistorical and historical compatibilists. Given Mele's contention that neither Beth nor Ann could do otherwise, it *appears* that we are to understand the case so that Ann's freedom in doing her philosophical work is derivative. If so, Ann's freedom is grounded in her freely making herself into this person. Prior (directly) free acts led her to acquire unsheddable values so strong that in later behavior those values rendered her unable to do otherwise. But then when Beth is manipulated into being a duplicate of Ann, she (Beth) is manipulated in such a way that she does not satisfy minimally sufficient conditions a plausible nonhistorical compatibilist theory would enlist for *directly* free actions. So understood, we cannot test for the adequacy of nonhistorical compatibilism with the case of Ann and Beth.

In what follows, I shall appropriate Mele's case of Ann and Beth to my ends. I ask the reader to understand the case so that both Ann and post-manipulation Beth possess and act from unsheddable values, but in which the values do not render either Ann or Beth unable to act otherwise. At the time of the acts in question, both satisfy conditions that nonhistorical compatibilists would take to be sufficient for directly free actions. Interpreted in this way, I believe the case is highly instructive.

⁴⁸ Mele (2006, pp. 164–166).

⁴⁹ McKenna (2004, p. 173).

⁵⁰ Mele (2006, p. 166).

⁵¹ See Mele (2008, p. 269, n13; and 2009b, pp. 464–465, and p. 466, n7).

⁵² That is, I had assumed *either* that Beth could have done otherwise, *or* instead that Beth was, say, suitably responsive to reasons, or something of that sort—something that nonhistorical compatibilists would make use of to account for directly free action.

In giving Beth unsheddable values from which she acts, we replicate a plausible psychological feature of actual human agents. As noted earlier (Sect. 1), I suspect that a considerable range of human action by most well-functioning actual persons shares this feature. We really are like that. And equally to the point, I also suspect that most well functioning human persons are such that, on many occasions, when they do act from unsheddable values, the causal role those values play is *not* so forceful that, at the time of action, these persons are literally unable to do otherwise (in some respectable compatibilist sense of that notion). While in no way do I wish to deny that this happens, I suspect it is rare. In any event, I turn to an assessment of the historical/nonhistorical debate in light of how I think it best to take the case of Ann and Beth.

5 In Defense of Nonhistorical Compatibilism: The Case of Suzie Instant

The reason I find the Ann and Beth case to be an especially useful when interpreted as I propose is precisely because it showcases the appeal of a plausible and philosophically interesting historical view. Both Ann and post-manipulation Beth, we may suppose, are equally capable of acting in response to norms of various sorts, or for that matter, in opposition to them. This we can stipulate is satisfied for each by virtue of the duplicate timeslice properties that they share. Both are therefore in this sense normatively competent agents. But only one, Ann, acts from a set of values—values helping to constitute her moral personality—that issue from and are integrated into her own history. These values, we can assume, were acquired non-coercively under her own steam through a process of critical reflection, evaluation, re-evaluation, modification, endorsement, and so on. When Beth acts, she acts from values that were simply forced upon her; her moral personality was settled for her, and her prior one was deleted, so to speak. That difference does, it seems to me, weigh heavily in favor of judging *only* Ann to act freely and be morally responsible for her acts.⁵³

In light of all of this, what can be said on behalf of the nonhistorical compatibilist? In order to lend support, in an earlier piece, I called upon the aid of a magical agent, Suzie Instant.⁵⁴ Suzie Instant is created by a god at an instant and is placed in a determined world. She is created to be a psychologically healthy woman indistinguishable from any other normally functioning thirty-year-old person whom any of us might encounter. To get this result, she is given a huge set of beliefs according to which she has lived a normal human life for 30 years. For instance, she believes (falsely) that she had a twelfth birthday and that her daddy bought her a pony. Furthermore, Suzie has some range of unsheddable values. She also has a set

⁵³ More recently, Mele has taken up cases of this sort, cases in which a globally manipulated agent has installed unsheddable values from which she acts, but in which it is nevertheless the case that this agent could have done otherwise (2009a, p. 173). In these cases, Mele is prepared to commend the (credible) historical compatibilist verdict that these agents are not morally responsible.

⁵⁴ In the remained of this paragraph and in the next two, I draw from McKenna (2004, pp. 180–181). In doing so, I have revised slightly the case of Suzie Instant so as to fit it for the points developed in this paper.

of false beliefs about how she came to acquire those values. She thinks that she acquired them through a process of sustained effort over the years leading up to what she thinks is her thirtieth. She takes pride in this fact and believes that she is responsible for this process and that she engaged in it freely (on this point, clearly she is mistaken). She is, also, a richly self-controlled person who is able to resist the inclination to act with weakness of will. Often when she acts, the desires issuing in her actions are the ones she wants to act upon, and when she does, she is sensitive to a wide range of reasons for action. Hence, Suzie satisfies an impressive set of features of the sort that, when she acts, varying compatibilists would regard those features as adequate for satisfying all of the nonhistorical conditions highlighted in their respective accounts of CAS.⁵⁵

Now suppose that Suzie is presented with the option to do one of two things, A or B. One option, B, involves a violation of a value that is unsheddable for her. The other option, A, involves acting from one of her unsheddable values. Suzie A-s, acting as her unsheddable value counsels, but in doing so, she could have done otherwise—that is, she could have B-ed. If, in the context of this debate, we are *already* operating under the assumption that compatibilism is true, it is not clear to me that Suzie did not act freely and that she is not responsible for what she does. It is difficult to see how a causal history that zeroed in on Suzie Instant all in an instant renders her unfree in a way that she would not be if instead some causal history or other unfolded over the course of 30 years. Note that when Suzie A-ed from her unsheddable value, she was *not* compelled to do so. Her doing so was *nothing like* acting upon an irresistible desire. It would be natural to say that she A-ed freely—in at least some non-question begging, restricted sense of freely, say freely*.⁵⁶

To press the point, imagine that every now and then this same god who created Suzie Instant visits another possible world and there creates another thirty-year-old Suzie, Suzie Normal, in the normal zygote manner. In this latter scenario, the god arranges things so that, after zygote-creation, little itsy-bitsy developing Suzie Normal is left alone to develop in the normal manner for a full 30 years (and 9 months). As it turns out, Suzie Normal at the age of thirty arrives at the precise point where she comes to be a nonhistorical qualitative duplicate of Suzie Instant. Suzie Normal faces the exact same choice between options A and B as Suzie Instant faces. Just like Suzie Instant, Suzie Normal opts to do A. In the case of Suzie Normal, it seems that the compatibilist—any compatibilist—will be hard pressed to

⁵⁵ For an excellent treatment of agents like Suzie Instant, see David Zimmerman’s paper, “Born Yesterday: Personal Autonomy for Agents without a Past” (1999). Mele comments that I, “seemingly take it to be clear that Suzie Instant is a conceptually possible agent” (2009b, p. 467). For the record, I do think such beings are conceptually possible, despite Davidson’s worries about swampman (Davidson 1987). I am not sure I would say I find it clear. For Mele’s position on this issue, see Mele (1995, pp. 172–173; and 2009a, pp. 174–179).

⁵⁶ One potential source of concern about this example is that when Suzie Instant A-s, she does so while non-culpably believing about herself many things that are false. Given reasonable epistemic constraints on moral responsibility, this might excuse or exempt her. But as Haji and Cuypers have pointed out (2007: 349–350), all that is required to avoid this pitfall is that Suzie’s act of A-ing be one that does not implicate any objectionably false beliefs.

deny that Suzie Normal A-s freely and is morally responsible for doing so.⁵⁷ But if so, how is it that Suzie Instant is rendered not free and not morally responsible when she A's at the relevant time merely by virtue of the fact that the causal history giving rise to her action came compressed in a momentary package where Suzie Normal's history chugged along over the course of 30 years? A difference here seems arbitrary.

While the case of Beth lends support to the historical compatibilist, I believe that the case of Suzie Instant lends support to the nonhistorical compatibilists. If so, how are we to proceed? To help bring the conflict into focus, I arranged it so that the Ann and Beth cases converged on the Suzie Normal and Suzie Instant cases.⁵⁸ To this end, imagine that merely by cosmic accident, not even by design, the god who created Suzie Instant and Suzie Normal happened to bring Suzie Normal into existence in such a way as to live out a duplicate life to the one lived by Ann, the only difference being that Ann's parents generated the initial zygote-later-to become-Ann in the typical and, dare I say, more pleasurable fashion. And suppose that Ann, after 30 years, comes upon the same choice that Suzie Normal and Suzie Instant faced between A and B, with the same unsheddable values held fixed, and so on. Imagine, for instance, that the options A and B are between continuing to work on an article through the weekend (A) and instead attending a beloved sister's wedding (B). In this case, as I see it, we should treat the case of Ann and Suzie Normal in the same way. Both A rather than B, and it is plausible to think, granting that compatibilism is true, that both A freely and are morally responsible for doing so. Now inject Suzie Instant into the case. If indeed we should treat her no differently than we treat Suzie Normal, then we should treat Suzie Instant no differently than we treat Ann. But, we may suppose, when Ann A-s rather than B-s, she can be treated as a nonhistorical qualitative duplicate of Beth in a context in which Beth also A-s rather than B-s. And so, finally, we can ask whether we should treat the case of Beth when she A-s no differently than we treat the case of Suzie Instant when she A-s.

Careful readers will note at this point that I have *not* claimed that we *should* treat the case of Beth and Suzie Instant in the same way, nor did I in my earlier formulation.⁵⁹ Nevertheless, it seemed to me then, and it still seems to me now, to be useful to consider how we ought to proceed under the initially plausible assumption that we should treat the cases symmetrically. Given this background assumption, should we use the leverage established with our reaction to Suzie Instant to force a similar treatment of Beth? Or should we instead use the leverage established with our reaction to Beth to force a similar treatment of Suzie Instant?

For those interested in what reason there might be for treating the cases of Beth and Suzie Instant in the same way, an argument to that effect might go as follows: Suzie Instant is alleged to be free and morally responsible on compatibilist grounds.

⁵⁷ The case of Suzie Normal is very much like the case of Mele's Ernie, who was created by the god Diana in zygote form and then set free to live out a normal human life (2006, pp. 188–189). According to Mele (2006, p. 193), and I agree, a compatibilist is committed to denying that when Ernie performs some act that Diana intended for him, he does not do so freely and is not morally responsible.

⁵⁸ McKenna (2004, pp. 181–182).

⁵⁹ McKenna (2004, p. 182).

Assume she is. If so, it at least *appears* that, solely by virtue of nonhistorical properties just subsequent to her A-ing, she satisfies sufficient conditions for A-ing freely and being responsible for it.⁶⁰ But, by hypothesis, Beth satisfies the same nonhistorical properties as does Suzie Instant when Beth A-s. Hence, if Suzie Instant A-s freely and is morally responsible for doing so, then Beth A-s freely and is also morally responsible for doing so. If, on the other hand, Beth does not A freely and is not morally responsible, it seems that the same ruling applies to Suzie Instant, since the case of Beth shows that the nonhistorical properties both satisfy are insufficient for acting freely.⁶¹

6 Treating the Cases of Beth and Suzie Instant Symmetrically

Grant for the moment that we should treat the case of Suzie Instant and the case of Beth in the same manner. Under this assumption, I have offered three considerations nonhistorical compatibilists like Frankfurt might take on in order to explain away the intuitive force of the Beth case⁶²:

First, Ann, we can stipulate, freely acquired her unsheddable values and thus is morally responsible for the acquisition of the moral personality from which she deliberates. The same point applies to Suzie Normal. This is not the case for either Beth or Suzie Instant. Perhaps part of the intuitive unease in holding that Beth is just as responsible *for A-ing* as Ann is, can be accounted for by attending to the legitimate concern that there should be some difference in judgments about Ann and Beth. But this much the nonhistorical compatibilist can accommodate, since the difference need not be explained in terms of one being free and morally responsible for A-ing and the other not, but rather in terms of one being morally responsible for more than the other.⁶³

Second, as Nomy Arpaly has pointed out, Beth's autonomy *has* been violated in one sense of that term.⁶⁴ She has been done wrong in a way that has robbed her of the moral personality that she fashioned for herself prior to the manipulation that was forced upon her. Perhaps part of our reluctance to treat Beth as freely A-ing and morally responsible for doing so is that we wrongly think that in making such a

⁶⁰ Below (Sect. 7), I will show how one might object that there is a further negative historical condition at work here.

⁶¹ In my earlier formulation (2004, pp. 181–182), I did not make this line of argument explicit, but I can see from Mele's discussion that I should have:

McKenna apparently is thinking that there is good reason to make symmetrical moral responsibility judgments about instant agents like Suzie Instant and radically manipulated agents like Beth. However, because he offers no argument for this thought, I cannot be blamed for my ignorance about why he finds it plausible. (Mele 2009b, p. 466)

I hope that what I have offered here helps. Note that in their critical discussion of my 2004, Haji and Cuypers (2007, p. 349) explicitly formulate and then discuss the line of argument I have set out here.

⁶² McKenna (2004, pp. 182–184).

⁶³ McKenna (2004, p. 183). Mele acknowledges and rejects this point, though he does not discuss it (2006, p. 172).

⁶⁴ Arpaly (2003, p. 128).

judgment, we are not recognizing the quite clear violations of Beth's rights as a person. But we can recognize that Beth freely and responsibly A-ed and *still* draw appropriate moral judgments about the moral wrongs done to Beth and how she deserves to be treated in light of that history. What, we might ask, could count as a proper moral response to Beth for her having suffered from someone else deciding for her what kind of person she should be? This question can be given a rich answer even if, now that she is this different (sort of) person, we are warranted in thinking that she is a person who acts freely and responsibly for what she now does.⁶⁵

Third, consider how *you* would think of and respond to Beth or Suzie Instant were you to have a morally charged transaction with one of them.⁶⁶ Imagine that you were aware of their unusual causal histories. Would you find it genuinely misplaced to make moral demands of them, to think it unfitting to respond with indignation or resentment were one of them to, say, insult your child or cunningly take illegitimate financial advantage of you? Perhaps at first blush it would seem misplaced. But think through the case as vividly as you might be able. Hold clearly in mind the point central to the nonhistorical compatibilist that, at the time coincident with the (putatively) free act, Beth is a nonhistorical duplicate of Ann. She has just as many resources for moral reasoning and deliberation as Ann does. She is just as able to do otherwise as Ann is. She adopts the same attitudinal stance toward her own convictions about her unsheddable values as Ann does (she is proud of how she believes herself to have acquire them, and so on). In such a case, I remarked, when vividly imagined, it can seem credible to claim that the right way to think of either Beth or Suzie Instant, and the right way to respond to either would be to treat either as a "real" person, one who is a fully competent moral agent and a legitimate target of our blaming responses. This is in the spirit of Frankfurt's remark that: "We are the sorts of persons we are; and it is what we are, rather than the history of our development, that counts. The fact that someone is a pig warrants treating him like a pig, unless there is reason to believe that in some important way he is a pig against his will and is not acting as he would really prefer."⁶⁷

To these three points previously developed, I would here add a *fourth*. Judgments of blameworthiness can be clouded by the delicate relationship between blameworthiness and the propriety or fittingness of blame, where the blame at issue is meant to be overt blame directed at the blameworthy party. On my view, when this sort of blame is fitting, in the sense of being deserved, it provides one who has standing to blame with a *pro tanto* reason to do so.⁶⁸ But that reason can be checked in the sense that it will not warrant an all-out reason in the face of certain defeaters. If, for instance, blaming someone for a minor peccadillo would cause her massive harm, or harm to innocent others, that could count as a defeater. Equally, if it would be too much of a strain on the one blaming, that too could count as a defeater. Furthermore, reasons can bear on mitigating blame or qualifying it. The child who lies might deserve blame, but the bullies harassing him into a state of

⁶⁵ McKenna (2004, pp. 183–184).

⁶⁶ McKenna (2004, pp. 182–183).

⁶⁷ Frankfurt (2002, pp. 27–28).

⁶⁸ McKenna (2011).

extreme anxiety might give one reason to focus more on a response of concern and go easy on any blaming. With this point in mind, revisit the point above about the kind of response Beth deserves as one whose previous moral personality had been hijacked. The legitimate response to her for that wrong done to her could give one good reason to go light on blaming her. Or it could give one good reason not to blame her at all. But it would not follow from this that she would not be *blameworthy* for her act of A-ing.

The point at issue here also bears on an especially provocative strategy of Mele's. In setting out cases like the Beth case, Mele invites his readers to imagine in Beth's place the sweetest person they know, such as one's grandmother.⁶⁹ Aside from the emotional noise this can generate as regards reliable intuitive responses, it is also true that for some to whom we are closely connected, we have good reasons not to blame them overtly, *even if they are blameworthy*. Grandmothers seem to me to be good candidates for being in that camp. There is, finally, the matter of practical reasons not to blame. Imagine beginning to blame someone who is clearly blameworthy just as she falls into a coma, or drops dead. The pointlessness of doing so would not change the fact that the person you were in the process of blaming was one who, just then, deserved blame; were she to have remained alive and conscious, she would have been an apt target of your blaming. This speaks to a further strategy Mele relies upon, an especially gripping one which involves immediate reversal of manipulation.⁷⁰ So, imagine that just after Beth's A-ing, the effects of the manipulation are reversed and she is returned to her former self. The natural response, it seems to me, is to think it would not be right to blame Beth. But the nonhistorical compatibilist can explain this away in terms of the reasons for blaming, not the reasons for judging Beth to be blameworthy at the time coincident with her action. Just like the person who dropped dead, or fell into a coma, so the nonhistorical theorist can argue, there is no point to blaming Beth. But for all of that, it simply does not follow that, just then, when Beth acted, she was not blameworthy. And were she to have persisted in that state, with that moral personality as it was at the time of her action, then it would have been fitting to blame her.

The four preceding points have not been offered as decisive grounds for a nonhistorical conclusion. All I have sought to establish, and all that the points enlisted above are designed to show, is that the nonhistorical compatibilist is not impotent in the face of the challenges at issue. A plausible case can be made for the view. It should not be regarded as a thesis so far off the charts that it ought to lead someone like Mele to write that "if compatibilists were to have nothing more attractive to offer than Frankfurt's—or any other—ahistorical view of moral responsibility and freedom, compatibilism would be in dire straits."⁷¹

I have thus far proceeded under the assumption that the cases of Beth and Suzie Instant should be treated symmetrically. Momentarily, I will explore grounds for resisting that assumption, but before doing so, I pause to note why the assumption is

⁶⁹ Mele (2009a, pp. 169–170; and 2009b, p. 465).

⁷⁰ For example, see Mele (2009a, p. 170).

⁷¹ Mele (2006, p. 172).

relatively benign for compatibilists committed to a positive historical thesis—that is, a thesis according to which it is a necessary condition for a directly free act that an agent have gone through a particular kind of freedom-and-responsibility-enabling history prior to so acting. Given a positive historical thesis, Beth's A-ing is unfree due to the fact that there was a particular kind of history that she lacked. She did not go through the process of taking on her newly acquired unsheddable values in the proper way because they were installed for her through some deviant source. But of course, having virtually no history at all, Suzie Instant also lacked the relevant kind of history. She too did not take on her unsheddable values in the prescribed way, and so she too should also be regarded as unfree.⁷² If on the other hand it is granted that Suzie Instant does A freely, then her case alone already establishes that directly free acts for which an agent is morally responsible do not require any particular kind of history, and so it should not matter what Beth's history is, so long as she is a nonhistorical qualitative duplicate of Suzie Instant. Hence, it seems that for the positive historical compatibilists, the cases of Beth and Suzie Instant stand or fall together.

7 Treating the Cases of Beth and Suzie Instant Asymmetrically

Here we come upon one reason why a negative historical thesis might be a better one in comparison with a positive thesis. A negative historical condition appears to license differential treatments of the cases of Beth and Suzie Instant.⁷³ And it licenses those differences in a manner that seems to accommodate most naturally the differing intuitions about the cases. If what a negative historical thesis requires is lacking a history of a certain sort, one that *is* freedom-and-responsibility-destroying with respect to a basically free action, then a negative historical compatibilist is able to argue that Beth does not freely A. This is because she does not satisfy the negative condition of lacking a freedom-and-responsibility-destroying history with

⁷² Thus, it is easy to see that Fischer and Ravizza (1998) are as committed to Suzie Instant's A-ing unfreely as they are to Beth's A-ing unfreely.

⁷³ One central criticism Mele leveled against my earlier discussion of his view (McKenna 2004) is precisely that he is on record as granting that some instant agents can be regarded as free and morally responsible for their actions while agents like Beth are not (Mele 2009b). Mele asks why he is under pressure to treat the cases of Beth and Suzie Instant symmetrically. It is a fair question, and I mean to take it up, but only indirectly. The problem with speaking directly to the difference between us at this juncture is that we apparently are talking about very different kinds of cases. As I have already made clear above (Sect. 4), I do not mean for the Beth case to be understood so that her action when she A-s is one that is only derivatively free, if it is free at all; I mean for it to be understood as one in which what is in question is whether her action is directly free. It appears that Mele does not think about his case of Beth in the same way—insofar as he is not thinking of his case of Ann as one involving what I would call direct freedom. And as for the matter of the instant agents that Mele is prepared to regard as free and responsible (e.g., 1995: 172–173), these are agents whose abilities far exceed the abilities of normal human persons like ourselves. All of their values are sheddable. I have, by contrast, sought to test a nonhistorical compatibilist thesis by considering a case of an instant agent who, when she is brought into existence, can be regarded as a current-timeslice replica of an actual human person. This is a constraint that I believe is warranted by virtue of the dialectical assumption that early on (Sect. 1) I claimed for compatibilism. So, under the highly plausible assumption that actual adult human persons cannot simply shed each of their values at will, instant agents with only sheddable values will not do.

respect that the causal antecedents of A-ing. Her unsheddable values were coercively installed. But, having no history to speak of, Suzie Instant *does* lack a history whereby her unsheddable values were coercively installed. Lacking any history, she also lacks one that is freedom-and-responsibility-destroying. Hence, so long as she satisfies other pertinent and uncontested nonhistorical conditions for basically free actions, it can be granted that she acts freely and is morally responsible for doing so.⁷⁴

To drive home the point, recall that above (Sect. 5) I briefly sketched an argument for why it might be thought that the cases of Beth and Suzie Instant should be treated symmetrically. A key premise was this one: If Suzie Instant is free and morally responsible for A-ing, she is so solely by virtue of satisfying nonhistorical properties just subsequent to her action. It is open to the negative historical compatibilist to argue that this premise is false. If indeed, Suzie Instant did act freely, it is only by virtue of the nonhistorical properties she satisfied *in conjunction with a further negative historical property* that she also satisfied—that is, the property of not having a history of the objectionable sort. Beth, on the other hand, did not satisfy *all* of the crucial properties Suzie Instant did, since Beth failed to satisfy a negative historical property; she *did* have a history of the objectionable sort. So it is erroneous to conclude that we should treat the case of Beth the same way we should treat the case of Suzie Instant.

But now the question arises, how convincing is a *purely* negative historical condition? I will now argue that the motivation for such a thesis falls short. To begin, consider these two different ways of formulating a negative historical thesis:

First, an agent A-s freely and is morally responsible for doing so only if, with respect to the causal springs of her A-ing, she has a history that *does not* include the acquisition of any unsheddable values through means that bypassed her ability to critically acquire, assess and sustain them.

And:

Second, an agent A-s freely and is morally responsible for doing so only if, with respect to the causal springs of her A-ing, she *does not* have a history that includes the acquisition of any unsheddable values through means that bypassed her ability to critically acquire, assess and sustain them.

Notice the scope of the ‘does not’ in each formulation. In the first, the mention of history falls outside that scope; in the second, it falls inside it. The first formulation presupposes that an agent who A-s freely *has* a history but that it does not include elements that are freedom-and-responsibility-destroying with respect to a candidate free act. The second formulation does not presuppose that an agent who A-s freely have any history at all. It is only the second that speaks directly to the case of instant agents like Suzie Instant. Here, I do not wish to claim that it is only the first that is plausible and not the second. Rather, I want to claim only that the plausibility of the

⁷⁴ Haji and Cuypers (2007) offer a different explanation of why the cases of Beth and Suzie Instant should be treated asymmetrically. They do so by way of defending a positive historical thesis, but one that only applies to agents who have a history. I will set aside their treatment here, as I have responded to them elsewhere (McKenna 2012).

second rests upon the prior plausibility of the first—not in the sense that the first is sufficient for the second, but just in the sense that any motive for embracing the second would be lacking if there were not good reason to embrace the first. And now, what I am interested in is what would underwrite the plausibility of the first.

Consider two ways of telling the story of Ann's history. The first, we can assume, is how we ought to have taken the case in the previous telling of it. In this version, Ann freely engaged in activity earlier in her life that led to the solidifying and perfecting of her unsheddable values. When in her undergraduate and graduate years she felt the call of the wild, enticing her to party the night away, she would sometimes fall prey to temptation. She would do such wild things as attend the weddings of her beloved family members rather than work on her writing projects. And at times, she would even do things like cultivate romantic relationships with reckless abandon. But she came to realize that these other things she valued stood in the way of her commitment to the life of the mind. So she worked hard to diminish the strength of these other values and inculcate the ones that she now cannot but possess with the impressive (though not overpowering) force they have. The second way of telling the story involves Ann*. Ann*'s parents were both obsessive professors. From an early age, Ann* passively followed them through life, mirroring their values and uncritically accepting their judgments about what is true and good. Ann* never once seriously considered the values she inherited from her parents, but rather with no effort invested in the shape of her evaluative outlook at all, just fell ass-backwards into her worldview, and eventually her values became just as unsheddable for her as Ann's are. Whereas Ann was an active value acquirer, Ann* was no more than a value-sponge. The salient point to note here is that both Ann and Ann* satisfy the negative historical requirement currently under consideration. Ann* was no more coerced into the acquisition of her unsheddable values than was Ann—and this is in stark contrast to the case of Beth.

If we were to regard as freedom-and-responsibility undermining an intervention into Ann*'s psychic life whereby very different unsheddable values were covertly forced upon her, ones upon which she subsequently acted, it would *not* be because Ann* was robbed of a moral personality that she came to possess under her own steam. This sort of rationale might be available for the case of Ann, but not for Ann*. Ann* did not acquire her character so formed under her own steam; she was passive with respect to the acquisition of it in the history of her own life. Why then should such intervention be regarded as objectionable as it bears on Ann*'s acting from a set of coercively installed unsheddable values? Unlike Ann, Ann* is not robbed of anything in which she has made an investment; having allowed the vagaries of life to buffet her about as they did, the acquisition of her values was just as happenstance as were the values covertly and coercively installed in Beth. Here, it seems to me, a very plausible rationale for why one would take such an intervention into Ann*'s life as objectionable as it bears on her free agency is the following: Ann* would thereby be robbed of the opportunity to fashion for herself her own values, and so, her own moral personality by virtue of abilities that she possessed. Granted, she made no effort to fashion her own evaluative standpoint, but then, she is still accountable for the person she has become—because, by omitting to engage in any active way, she allowed herself to arrive at the person that she now is.

In sum, what is so important about an agent's having a history that lacks the acquisition of pertinent values through means bypassing her ability to critically assess them is that she thereby *has* a history that afforded her an opportunity to shape her moral personality for herself. And this can be so even if she never took advantage of that opportunity.

I find the proposed rationale to be a fairly attractive one for an historical thesis. But notice that it invites a positive historical thesis as the deeper one underwriting the credibility of the first of the two negative proposals currently under consideration. What is required with respect to the causal springs of an agent's putatively free act is that she *had* a history in which she had the opportunity to fashion her evaluative standpoint for herself from important agential abilities available to her, and (here is the negative ingredient) none of her unsheddable values were acquired by any means that bypassed her ability to critically, acquire, assess and sustain them. If this is the deeper historical rationale at work in accounting for why it would be wrong to intervene in the case of Ann*, and why it was wrong to intervene in the case of Beth, it is *also* a rationale for why Suzie Instant ought also to be regarded as unfree for her act of A-ing. If so, we do not after all have a basis for asymmetrical judgments as between the cases of Beth and Suzie Instant.

8 The Advantages and Disadvantages of Nonhistorical Compatibilism

Supposing, then, that there *is* pressure to treat the cases of Beth and Suzie Instant symmetrically, how ought we to proceed? I have granted that the case of Beth lends powerful intuitive support in favor of an historical conclusion, which, under the supposition of symmetrical treatment, would yield the verdict that Suzie Instant *does not* A freely. On the other hand, I have argued that the case of Suzie Instant lends intuitive support in favor of a nonhistorical treatment, which under the supposition of symmetrical treatment, would yield the verdict that Beth *does* A freely. The nonhistorical compatibilists, I argued (Sect. 6), have resources available to them to help explain away the appearance that the case of Beth clearly commands an historical conclusion. Nevertheless, I have in no way claimed that these resources are decisive and that nonhistorical compatibilism is the correct view. While I have been essaying on behalf of nonhistorical compatibilism, I wish to remain agnostic. Cases such as the case of Beth, as I have framed it, continue to strike me as powerful ones that encourage an historical conclusion. All I have sought to show by exploring a case like that of Suzie Instant, and by aligning it, so to speak, with the case of Beth, is that nonhistorical compatibilism remains a credible thesis. It is not demonstrably refuted by the arguments historical compatibilists have leveled against it.

Admittedly, nonhistorical compatibilism comes with a considerable cost that historical compatibilism does not, as I made clear early on (Sect. 2). Inject a case like the case of Beth into the template of the manipulation argument, and nonhistorical compatibilists are forced to take a hard-line to that particular argument by denying the premise that manipulated Beth is not morally responsible. The historical compatibilists, by contrast, can grant this premise—which, no doubt, the incompatibilist will contend is highly plausible—and instead adopt a soft-line by

arguing that a freedom-enabling deterministic history is relevantly different from a history like that of Beth's.

But this disadvantage weighed in the scales, it should also be noted that historical compatibilism is open to certain liabilities as well. There is, for one thing, the general worry that inviting historical considerations about the origins of agency plays to the hand of incompatibilists.⁷⁵ There is also the possibility that proposed necessary historical conditions will be open to counterexample. Fischer and Ravizza, for instance, require that an agent is morally responsible for what she does only if she adopts a certain subjective attitude toward her own agency, roughly, by believing that she is a morally responsible agent.⁷⁶ But, as various critics have noted, there are gripping cases that can be generated whereby an agent who does not take herself to be accountable to others is nevertheless a morally responsible agent.⁷⁷ Furthermore, there is the prospect of engendering a distinct source of skeptical concern—one that does not have to do with the general worry that universal determinism is true. Consider, for instance, the relatively lean positive historical proposal floated just above (Sect. 7):

What is required with respect to the causal springs of an agent's free act is that she *had* a history which afforded her the opportunity to fashion her evaluative standpoint for herself from agential abilities she possessed, and none of her unsheddable values were acquired by any means that bypassed her ability to critically, acquire, assess and sustain them.

I find this historical proposal extremely appealing. It seems to capture something important about our conception of freedom and moral responsibility, and perhaps autonomy too. Regardless, it is open to the worry that the range of actual persons who are morally responsible for a considerable swath of their conduct is *much* narrower than it is taken to be, and for two reasons. First, it is just not clear how many persons do have the opportunity to assess for themselves the values from which they act. Second, even those who do, often do so from highly limited epistemic resources. As Manuel Vargas has forcefully argued, a person's best efforts at character formation might yield results far different from what anyone could reasonably expect.⁷⁸ When later they come to possess and act from certain unsheddable values, often there is little credibility to claiming that their freedom and responsibility is to be accounted for in part by these earlier gambles. "How were they to know?" so the thought would go.

Of course, one advantage of nonhistorical compatibilism is just that it does not shoulder these potential disadvantages of an historical thesis. But another, more compelling one is just that it gets its traction from a very basic moral insight—that once a person has come to be a mature adult, that is, a morally responsible agent, she ought to be evaluated for what she does and how she is, not how she came to be that way.

⁷⁵ See, for example, Double (1991), and Watson (1999).

⁷⁶ Fischer and Ravizza (1998, pp. 220–223).

⁷⁷ See Eshleman (2001), Mele (2000), and McKenna (2000). For Fischer's reply, see his (2006).

⁷⁸ Vargas (2005). For a similar point, see Adams (1985), and McKenna (2008b).

9 Concluding Remarks

How does my treatment of the historical/nonhistorical debate fit with my contention that, where possible, compatibilists should favor a hard-line reply to manipulation arguments over a soft-line reply?⁷⁹ Given how I have framed the historical/nonhistorical debate, does this not mean that on my view compatibilists should favor nonhistorical compatibilism? Not at all. My advice for compatibilists with respect to favoring a hard-line reply to manipulation arguments is simply the product of two factors.⁸⁰ First, as I see it, for any world at which determinism is true, it is at least in principle a metaphysical possibility that, whatever the causal springs of an agent's actions happen to be, there is a way to replicate them by artificial means. So, eventually, incompatibilists will be able to come up with *some* manipulation story that gets the causal details “just right” as would be required by whatever credible formulation of CAS (Compatibilist-friendly Agential Structure) anyone could cook up. Second, when an incompatibilist has on offer a manipulation argument that does not quite get the details of a compatibilist's preferred account of CAS just right, *of course* the compatibilist has the option of taking a soft-line reply rather than a hard-line; that is, it is always open to her, in terms of the dialectic, to reply by pointing out that the manner of manipulation is relevantly different from a mere deterministic history. But, I contend, the deeper game cannot be won that way. It will always be open to an incompatibilist to counter by revising her example so as to satisfy her compatibilist opponent.⁸¹ Hence, when an incompatibilist comes close with a powerful manipulation case, it is better for the compatibilist to respond by taking on the example directly: revise the example as needed to satisfy CAS fully (whatever CAS is taken to amount to), and then just adopt the hard line.

This advice to compatibilists should *not* be taken to drive the debate between historical and nonhistorical compatibilists. Advising that compatibilists adopt a general policy of favoring a hard-line reply when possible is not to say that they should always dogmatically do so. If free action and moral responsibility really are historical notions, so be it. If they are, then any manipulation case/argument that severely violates the relevant historical constraints will fall short, and a compatibilist would be well-advised to take on a soft-line reply. In one sense, then, my treatment of the historical/nonhistorical debate is an entirely separate matter from my position regarding the way that compatibilists ought to respond to manipulation arguments.

Nevertheless, it is certainly true that there is an interesting relationship between these two debates. Manipulation cases like that of Beth, or, for that matter, Suzie Instant, are far more outrageous, and are for this reason much further from ordinary, every-day cases than are manipulation cases that have built right into them conditions of a sort that historical compatibilists would endorse. The closer the

⁷⁹ McKenna (2008a).

⁸⁰ McKenna (2008a, p. 144).

⁸¹ In fact, one can find a nice illustration of this point in Pereboom's formulation of his manipulation argument. In response to Fischer and Ravizza's historical conditions, which they contend render them immune to certain manipulation worries (Fischer and Ravizza 1998, pp. 235–236), Pereboom just proposes that their historical requirements be added to his manipulation case (Pereboom 2001, p. 122).

incompatibilist's proposed manipulation story gets to real-life cases with all of the relevant history loaded in, the less intuitively jarring it is for the compatibilist to adopt a hard-line reply.⁸² So, if in the end historical compatibilism is true, this does make it easier for compatibilists to respond to manipulation arguments.

In closing, I will offer a few speculations as to why I apparently cannot come off the fence as regards the historical/nonhistorical debate.⁸³ It might well be that our concept of moral responsibility is vague or otherwise indeterminate in some way. Or instead, it might be that there just is no one single concept of moral responsibility. Cases like the Suzie Instant case pull in one direction, while cases like the case of Beth pull in the other. Thinking about the case of Suzie Instant helps to bring forth a conception of moral responsibility in which what matters is what a person does and how she is, regardless of her history. Thinking about the case of Beth helps to bring forth a conception of moral responsibility in which a person's responsibility for what she does and who she is has its roots at least partially in the role she has played (or at least had the chance to play) in fashioning her moral character for herself. The former is in a way more elegant, and it seems to accept in a more forthright fashion the hard truth all compatibilists must in the end take on—that, as Harry Frankfurt might put it, we are fashioned and sustained by factors over which we have no control.⁸⁴ Nevertheless, I must admit that these nonhistorical currents in our thinking about freedom and moral responsibility also seem to me to be by contrast with the historical conception more shallow; the notions of freedom and moral responsibility they commend turn out to be less important, less worth wanting. I am, however, at a loss to explain why exactly it is that I take this to be so.

References

- Adams, Robert Merrihew. 1985. Involuntary sins. *The Philosophical Review* 94: 3–31.
- Arpaly, Nomy. 2003. *Unprincipled virtue*. New York: Oxford.
- Arpaly, Nomy. 2006. *Merit, meaning and human bondage*. Princeton, NJ: Princeton University Press.
- Ayer, A.J. 1954. Freedom and necessity. In *Philosophical essays*, 3–20. New York: St. Martin's Press.
- Berofsky, Bernard. 1987. *Freedom from necessity*. London: Routledge & Kegan Paul.
- Berofsky, Bernard. 2006. Global control and freedom. *Philosophical Studies* 131: 419–445.
- Brand, Myles, and D. Walton (eds.). 1976. *Action theory: Proceedings of the Winnipeg conference on human action*. Dordrecht: D. Reidel.

⁸² This point is most clearly on display with Mele's forceful original design argument for incompatibilism (2006, pp. 188–189). This argument features the goddess Diana, who creates Ernie with the intention that he performs some particular "free" act at a certain point in his adult life, which Ernie subsequently does. Construed as a kind of manipulation argument (manipulating Ernie from a point in time prior to his coming into being), this example allows for a case in which Ernie's entire life history prior to his "free" act need not differ in any way from any other agent at a deterministic world. Aside from his manner of creation, Ernie really does live out the kind of ordinary, every-day life any other person might. Here the intuitive unease of claiming that Ernie does act freely and is morally responsible is ratcheted way down as in comparison with how jarring a case like the case of Beth is. It is little wonder that Mele rightly remarks that compatibilists should take (what I would term) a hard-line reply to this argument (2006, p. 193).

⁸³ I am indebted to Terry Horgan for this suggestion.

⁸⁴ Frankfurt (2002, p. 28).

- Buss, Sarah, and Lee Overton (eds.). 2002. *Contours of agency: Essays on themes from Harry Frankfurt*. Cambridge: The MIT Press.
- Campbell, Joseph. 1997. A compatibilist theory of alternative possibilities. *Philosophical Studies* 88: 319–330.
- Davidson, Donald. 1987. Knowing one's own mind. *Proceedings and Addresses of the American Philosophical Association* 60: 441–458.
- Demetriou, Kristin. 2010. The soft-line solution to Pereboom's four-case argument. *Australasian Journal of Philosophy* 88(4): 595–617.
- Double, Richard. 1991. *The non-reality of free will*. Oxford: Oxford University Press.
- Dworkin, Gerald. 1988. *The theory and practice of autonomy*. Cambridge: Cambridge University Press.
- Eshleman, Andrew S. 2001. Being is not believing: Fischer and Ravizza on taking responsibility. *Australasian Journal of Philosophy* 79: 479–490.
- Fischer, John Martin. 2006. The free will revolution (continued). *The Journal of Ethics* 10: 315–345.
- Fischer, John Martin. 2002. Frankfurt style compatibilism. In eds. Sarah Buss and Lee Overton, 1–26.
- Fischer, John Martin, and Mark Ravizza. 1998. *Responsibility and control: An essay on moral responsibility*. Cambridge: Cambridge University Press.
- Frankfurt, Harry. 1971. Freedom of the will and the concept of a person. *The Journal of Philosophy* 68: 5–20.
- Frankfurt, Harry. 1975. Three concepts of free action II. *Proceedings of the Aristotelian Society* Supp. Vol. II: 113–125.
- Frankfurt, Harry. 1988. *The importance of what we care about*. Cambridge: Cambridge University Press.
- Frankfurt, Harry. 2002. Reply to John Martin Fischer. In eds. Sarah Buss and Lee Overton, 27–31.
- Gert, Bernard, and Tim Duggan. 1979. Free will as the ability to will. *Nous* 13: 197–217.
- Haji, Ishtiyaque. 1998. *Moral appraisability*. Oxford: Oxford University Press.
- Haji, Ishtiyaque. 2009. *Incompatibilism's allure*. Peterborough: Broadview Press.
- Haji, Ishtiyaque, and Stefaan Cuypers. 2004. Responsibility and the problem of manipulation reconsidered. *International Journal of Philosophical Studies* 12: 439–464.
- Haji, Ishtiyaque, and Stefaan Cuypers. 2007. Magical agents, global induction, and the internalism/externalism debate. *Australasian Journal of Philosophy* 85(3): 343–371.
- Horgan, Terry. 1985. Compatibilism and the consequence argument. *Philosophical Studies* 47: 339–356.
- Kane, Robert. 1996. *The significance of free will*. Oxford: Oxford University Press.
- Kearns, Stephen. 2011. Aborting the zygote argument. *Philosophical Studies* (online April).
- Lehrer, Keith. 1976. 'Can' in theory and practice: A possible worlds analysis. In eds. Brand Myles and D. Walton, 241–270.
- Lewis, David. 1981. Are we free to break the laws? *Theoria* 47: 113–121.
- Locke, Don. 1975. Three concepts of free action I. *Proceedings of Aristotelian Society* Supp. Vol. II: 95–112.
- McKenna, Michael. 2000. Assessing reasons-responsive compatibilism. *International Journal of Philosophical Studies* 8: 89–114.
- McKenna, Michael. 2004. Responsibility and globally manipulated agents. *Philosophical Topics* 32: 169–182.
- McKenna, Michael. 2008a. A hard-line reply to Pereboom's four-case argument. *Philosophy and Phenomenological Research* 77: 142–159.
- McKenna, Michael. 2008b. Putting the lie on the control condition for moral responsibility. *Philosophical Studies* 139: 29–37.
- McKenna, Michael. 2011. *Conversation and responsibility*. Oxford: Oxford University Press.
- McKenna, Michael. (2012). Defending the case for nonhistorical compatibilism: A reply to Haji and Cuypers. *Philosophical Issues* 22.
- Mele, Alfred. 1995. *Autonomous agents*. Oxford: Oxford University Press.
- Mele, Alfred. 2000. Reactive attitudes, reactivity, and omissions. *Philosophy and Phenomenological Research* 61: 447–452.
- Mele, Alfred. 2006. *Free will and luck*. Oxford: Oxford University Press.
- Mele, Alfred. 2008. Manipulation, compatibilism, and moral responsibility. *The Journal of Ethics* 12(304): 263–286.
- Mele, Alfred. 2009a. Moral responsibility and agents' histories. *Philosophical Studies* 142: 161–181.
- Mele, Alfred. 2009b. Moral responsibility and history revisited. *Ethical Theory and Moral Practice* 12: 463–475.

-
- Pereboom, Derk. 2001. *Living without free will*. Cambridge: Cambridge University Press.
- Pereboom, Derk. 2008. A hard-line reply to the multiple-case manipulation argument. *Philosophy and Phenomenological Research* 77: 160–170.
- Schoeman, Ferdinand. 1987. *Responsibility, character, and the emotions: New essays in moral psychology*. Cambridge: Cambridge University Press.
- Smith, Michael. 2003. Rational capacities, or: How to distinguish recklessness, weakness, and compulsion. In eds. Sarah Stroud and Christine Tappolet, 17–38.
- Stroud, Sarah, and Christine Tappolet (eds.). 2003. *Weakness of will and practical irrationality*. Oxford: Oxford University Press.
- Taylor, Richard. 1974. *Metaphysics*. Englewood Cliffs: Prentice Hall.
- van Inwagen, Peter. 1983. *An essay on free will*. Oxford: Clarendon Press.
- Vargas, Manuel. 2005. The trouble with tracing. *Midwest Studies in Philosophy* 29: 269–291.
- Vargas, Manuel. 2006. On the importance of history for responsible agency. *Philosophical Studies* 127: 351–382.
- Vihvelin, Kadri. 2004. Free will demystified: A dispositional account. *Philosophical Topics* 32: 427–450.
- Wallace, R.Jay. 1994. *Responsibility and the moral sentiments*. Cambridge: Harvard University Press.
- Watson, Gary. 1999. Soft libertarianism and hard compatibilism. *The Journal of Ethics* 3(4): 351–365.
- Watson, Gary. 2002. Volitional necessities. In eds. Sarah Buss and Lee Overton.
- Wolf, Susan. 1987. Sanity and the metaphysics of responsibility. In ed. Schoeman.
- Zimmerman, David. 1999. Born yesterday: Personal autonomy for agents without a past. *Midwest Studies in Philosophy* 23: 236–266.