

ON THE IMPORTANCE OF HISTORY FOR RESPONSIBLE AGENCY

Manuel Vargas

(Forthcoming in *Philosophical Studies*)

One effective way of measuring the power of a philosophical proposal is to test it against a problem where existing proposals are at loggerheads. This is the broad aim of this paper. The proposal I consider here is a *moderate revisionist* account of responsible agency. As a first approximation, moderate revisionism is the view that a plausible and normatively adequate theory of responsibility will require principled departures from commonsense thinking. The problem I will consider is whether morally responsible agency— for my purpose, something that is settled by whether an agent is an apt target of our responsibility-characteristic practices and attitudes— is an essentially historical notion. Some have maintained that responsible agents must have particular sorts of histories, others have argued that no such history is required. Resolution of this contentious issue is connected to a wide range of concerns, including the significance and culpability of different forms of manipulation, the plausibility of important incompatibilist criticisms of compatibilism, and of course, a satisfactory account of moral responsibility.

I argue that my moderate revisionist account is among the existing proposals uniquely well suited to explaining the apparent intractability of the issue, and it provides the most plausible answer to the question of the importance of history for responsible agency. As it turns out, history matters sometimes, but less frequently than we might think.

1.

One assumption of what follows is that commonsense thinking about responsibility has commitments that are incompatible in some deep way with determinism being true.¹ I depart from most incompatibilists in thinking it is unlikely that we will find our self-image borne out by an independently plausible picture of the world. I will not try to defend these claims in this paper.²

Instead, I will take this collection of views, what I call a “folk conceptual (probable) error theory”, as a provisional starting point. Even so, I do not think, as traditional hard determinists tend to think, that it follows from a folk conceptual error theory that there is no moral responsibility, and that responsibility-characteristic practices and attitudes (praise, blame, etc.) are not justified. A sizable subset of our responsibility-characteristic practices, attitudes, and beliefs can be justified irrespective of the folk conceptual error, indeed, with greater epistemic and normative confidence than before. This paper constitutes one part of the argument for that conclusion.³

Revisionism has two aspects: diagnostic (the folk conceptual error theory) and prescriptive (the subsequent proposal for revision). The version I propose is distinct from standard forms of compatibilism in its diagnosis: moderate revisionism holds that commonsense thinking about moral responsibility is incompatibilist in some deep way. It is distinguished from standard incompatibilist theories in virtue of holding that the truth of determinism would do little to undermine the normative integrity of the bulk of our responsibility practices. Though the diagnosis it gives departs from mainstream theorizing about responsibility, it incurs several advantages. First, it need not deny the persuasiveness of arguments for the incompatibility of determinism and responsibility (as we ordinarily understand it). Second, it need not be committed to the troublesome metaphysics of libertarianism, or the drastic conclusions of hard determinism. Whether the prescriptive component of it carries similar advantages is something we will evaluate of the course of this paper.

The primary focus of this paper concerns revisionism about the concept of moral responsibility. Other foci are possible, as our concern with responsibility is bound up with at least three interrelated but distinct elements, any of which may be revisable. First, there are the characteristic psychological dispositions or attitudes associated with responsibility (roughly, what Strawson called the “reactive emotions,” e.g., resentment and gratitude). Second, there are the practices associated with responsibility (e.g., sanctioning, rewarding, and on some views, praising,

and blaming). Finally, there is the cluster of beliefs, judgments, or concepts about responsibility (including beliefs about the conditions for appropriate ascription of responsibility) that we might broadly call the folk concept of responsibility. The revisions I am presently concerned with are, at least initially, largely conceptual.

What makes moderate revisionism moderate is that it prescribes neither the eliminativism that we can call *strong revisionism* (e.g., skepticism about moral responsibility) nor the trivial or *weak* revisionism implicit in views that hold that while we may say one thing, we really mean, intend, or have in mind something else.⁴ Instead, it argues for significant but not radical revisions on the basis of two standards. The first is a standard of naturalistic plausibility. That is, any proposed revision should be constrained by a scientifically plausible conception of the world. Other things equal, the fewer demands a theory places on the outcome of current or future science, the better, and the more it coheres with what we have reason to believe, the better. The second standard is one of normative adequacy. That is, the revised concept or theory has to provide justified normative support for a sizable subset of our responsibility-characteristic practices, attitudes, and beliefs in a way that is well integrated with other justified norms.⁵

Revisionism, even in the moderate form I have outlined, can raise the worry that we are doing ersatz problem solving by changing the topic from responsibility to something else. Though the semantic issues here are complex, we need not worry. The point of a theory of responsibility is to tell us the conditions under which we are entitled to treat people in those ways characteristic of responsibility. Correspondingly, my concern is whether the bulk of our responsibility-characteristic practices, attitudes, and beliefs, are justifiable in light of a plausible picture of agency. I think they are, and will attempt to show why. Whether or not we call the resultant account a theory of responsibility or something else is a secondary issue. Still, it seems to me that, to the extent that the proposed revision is conservative, preserving the bulk of responsibility-characteristic beliefs,

practices, and attitudes should be enough to earn the right to claim to be a revisionist theory of responsibility. Nonetheless, if this account can only be an account of responsibility*, and not responsibility proper, we still have reason to attend to it, as a theory of responsibility* is precisely a theory of the justified normative elements we associate with responsibility.⁶

Let us turn to the issue of the importance of history for responsible agency.

2.

We are presently concerned with morally responsible agency— accountability, if you like— and not with the degree of blame or praise a particular agent merits for some action or inaction. These two phenomena are connected: minimally, morally responsible agents are candidates for praise and blame. However, my focus is primarily on a particular aspect of responsible agency— whether responsible agency is, as John Martin Fischer and Mark Ravizza have put it, “essentially historical.”⁷

There are at least three views one can have about the importance of history for responsible agency.

First, a theory counts as *historical* if it holds that a responsible agent must have had a particular kind of history. The required history varies by account. In some cases, the history must satisfy a negative condition. For example, Alfred Mele has argued that an agent’s capacities must have been acquired in a compulsion-free way.⁸ Other theories require the satisfaction of a more demanding positive condition. Fischer and Ravizza argue that an agent's responsibility depends, in part, on the agent having "taken responsibility" for his or her "reasons-responsive mechanism" (1998, 210).⁹ Thus, history might be taken to matter in either (or both) negative or positive ways.

Second, a *structural* theory holds that history is unimportant, and that what matters for responsibility is whether the agent at the moment of action had the relevant psychological structure(s) in place. It is natural to read at least Harry Frankfurt's early work as committed to this

idea.¹⁰ On a simple "hierarchical theory," a suitable agent is morally responsible if and only if it has a desire to act upon the desire that is its will. In his classic example, an unwilling addict who gives in to her overpowering urge to take a drug could not be said to want the drug-taking desire to be her will, and thus, is not a responsible agent. By contrast, a willing addict counts as responsible because he both wants to take the drug and he wants the desire to take the drug to be his will. However, in neither case does the history of the agent matter for responsible agency. Rather, what matters is the occurrent desiderative structure of the agent. So, on this kind of view, what matters is the structure of agency, not its history.¹¹

Several clarifications are in order. A structural theorist need not claim that history is always irrelevant to the assignment of responsibility. That is, it may be that the only way we can learn about the structure and powers of an agent is by attending to the agent's history. So history can, as a practical matter about the epistemology of responsible agency, play a role in how we evaluate the capacities of an agent. However, the practical importance of history is ultimately parasitic on the ontological issue we are trying to settle. Similarly, we might find that the precise assignment of praise or blame for some action depends on certain historical features of the agent. But here too, this is a further issue that is parasitic on whether the agent is accountable in the first place. We cannot assume that the conditions for responsible agency will turn out to be the very same conditions that determine the degree of praise and blame.

We also ought not get confused by a further historical feature lurking in considerations about responsible agency. Some structural features of agency (deliberation, for one) have a certain amount of temporal extendedness. The sort of history— history in the process of deliberating or being responsive to reasons— is not what separates the structuralist from the historicist. Rather, it is the issue of whether or not a particular pre-deliberative or pre-process-of-decision elements and/or processes require a particular history.¹²

We can distinguish a third kind of theory, something I will call a semi-structural theory, from structural and historical accounts. (If we were feeling especially cheeky, we might call it a "post-structural theory of agency.") One might hold that a particular kind of history is sufficient for moral responsibility without thinking that it is necessary for moral responsibility. That is, one could concede that history can matter, but still insist that there are cases of responsible agency where history is irrelevant.¹³

One reason to make a concession in the direction of history comes from so-called "tracing" cases. Oftentimes, we seem to trace responsibility for an action past the immediate structure of agency back to some earlier point in the agent's history. Drunk driving is one example. We hold someone responsible for the results of drunk driving not because of the kind of agent they are when they get behind the wheel, but rather, because of the kind of agent they were when they started to drink, knowing that drinking impairs judgment and coordination, and so on. That is, even if the agent does not satisfy the structural requirements for responsibility at the time of action, we can trace responsibility back to a prior state where the agent did satisfy the structural conditions in a way that generates responsibility for the later actions. So, even if the structure of agency is important, this does not rule out the need to acknowledge that a certain kind of history may be required for responsibility.¹⁴

The semi-structuralist attempts to vindicate the spirit of structuralism, but not the letter: history might be sufficient for responsibility in the absence of the relevant responsibility-supporting agential structures, but in cases where those structures are present, so is responsible agency. Though semi-structuralism is, at least to my mind, an advance over simple structuralism, there is reason to think history has more pull on our responsibility ascriptions than the semi-structuralist admits.

Consider the following case: Octavio is walking down the street, sees an attractive woman, and proceeds to whistle at her. Suppose we know that Octavio is free and responsible in Frankfurt's

sense. That is, his whistling issues forth from a desire to whistle, where he also has a second-order desire that his desire to whistle be his will. On either a structural or semi-structural theory of Frankfurtian character, these facts are enough to make Octavio a responsible agent. Now suppose, though, that we also learn that Octavio normally behaves like a perfect gentlemen, and even if he occasionally has the desire to whistle, he is always able to squelch it. But, suppose that we also learn that in this instance, his will had the particular structure it did because a nefarious neurosurgeon (of the sort that roams the free will literature) was somehow able to secretly implant the relevant desiderative structure into his mind immediately before the attractive woman walked by. On a simple Frankfurtian account, as this example presumes, the entire volitional structure could be imposed.¹⁵ Thus, we may say that Octavio comes to whistle in this instance because of the work of the manipulator.

Though intuitions can differ, it seems that the typical reaction to this case is to say that if Octavio's whistling was brought about by manipulation, then we ought not treat Octavio as a responsible agent, regardless of the condition of his agential structure. If so, it looks like history matters both in cases where the relevant structure is not present and in cases where the structure is present. In short, historicism, not semi-structuralism seems to be the right view.¹⁶

If this seems too pat, it is because it is. Structuralists and semi-structuralists will dig in at this point, arguing that if the manipulation leaves the agent's relevant agential structures intact (e.g., in Octavio's case, the second order desire that his effective first order desire be his will), then what is left is still a case of responsible agency. On the other hand, if the manipulation destroys or impairs those structures, then there is less reason to think that responsible agency has survived the manipulation.¹⁷ But, if Octavio is not responsible, it is for purely structural (i.e., non-historical) reasons.¹⁸

Several issues make analysis of these cases anything but straightforward. First, Michael Bratman has noted that our view of the nature of reasons may affect how we think about a post-manipulation agent.¹⁹ If we are externalists about normative reasons, then changing the structure of an agent so that the post-manipulation agent is strongly responsive to reasons might, he thinks, leave us inclined to think the agent is still a responsible agent because the ends of the agent are fixed by something external to the agent. In contrast, on an internalist picture of reasons where one's reasons are at least partially fixed by contingent features of the agent, there may be greater temptation to think that manipulation of the agent (and thus, her reasons) undermines responsibility even if the manipulated agent is responsive to her manipulated reasons. Second, it may also make a difference whether we say the relevant agential structure persists through pre- and post- manipulation agents or whether pre-manipulation agent's structure is replaced with another structure. Though establishing this with any certainty would surely require empirical investigation (e.g., controlled surveys of untutored subjects), it is reasonable to believe that people would be less likely to ascribe responsibility in cases where the relevant agential structure was replaced than in cases where the relevant agential structure was intact, but just subject to manipulated inputs (e.g., adding desires or values to an agent vs. tampering with the reasons-responsive mechanism more directly). Moreover, this is likely to interact with the considerations Bratman raises. A third variable is how one thinks of personal identity. More dramatic manipulations will raise concerns about the continuity of the person than less dramatic manipulations. These issues are all delicate and have complex interconnections, which should make us suspicious about whether there are easy lessons to be drawn from manipulation cases, especially via thought experiments alone. Until we carefully do the required empirical work, these issues are likely to render unreliable arguments that rely exclusively on manipulation cases.

Historicists might attempt to get around some of these difficulties with an indirect argument. Fischer and Ravizza have suggested one. They argue that tracing examples

at least show that there are certain contexts in which moral responsibility ascriptions do not supervene on responsiveness profiles, but depend crucially on history. This can make it at least plausible that *in general* such ascriptions depend on history, especially in light of the fact that it may seem odd that history matters only in certain cases (where there is a lack of responsiveness) but not in others (where there is responsiveness). On what basis could it be though that there would be this sort of asymmetry in the relevance of history?²⁰

These remarks suggest that unless we have a principled explanation for why we care about some history and not others, the semi-structural view appears to historicists as adding *ad hoc* epicycles to a troubled structural picture.

John Fischer introduced the useful term “dialectical stalemate” to refer to cases where well-intentioned but disagreeing parties lack the dialectical resources to move each other. This appears to be one of those cases. On the one hand, this form of indirect argument is likely to do little to change the mind of antecedently convinced semi-structuralists. Semi-structuralists can object that historicists’ insistence on symmetry in the importance of history is undermotivated.²¹ Moreover, there is no motivation to care about history when it plays no role in the decision of the agent.²² The fact that we sometimes care about history no more supports the plausibility that we always care about history than it supports the plausibility that we only sometimes care about history. So, semi-structuralists will remain unmoved by the arguments of historicists. On the other hand, Fischer and Ravizza are likely right that without a principled explanation of why we should care about some history and not others, it is unlikely that semi-structuralists will convince antecedently convinced historicists, either. Without a different sort of argument, we appear to be at loggerheads.

It is important to note a point on which some of this literature may be misleading; this stand-off does not obviously track a difference between compatibilist and incompatibilist intuitions. Several philosophers have suggested there is a conceptual pull between historicism and incompatibilism on the one side, and on the other, structuralism and compatibilism.²³ However, most compatibilists who have clear commitments on this issue seem to be historicists.²⁴ Indeed, most philosophers who have thought about the issue at all — incompatibilist or compatibilist— appear to be historicists. (Perhaps some will think this fact itself counts as evidence in favor of historicism!) However, it would be a mistake to assume that the history/structure disagreement is simply another facet of a more fundamental disagreement between compatibilists and incompatibilists.

For example, all the views about history I have canvassed either already exist or could be reproduced in incompatibilist terms. We might imagine Sartre or someone with an Epicurean-like view to be a structural libertarian. Laura Ekstrom's libertarianism includes a negative historical condition that, like Mele's account, makes her a historicist.²⁵ And, at least one formulation of Galen Strawson's pessimistic incompatibilism seems to be semi-structuralist; he maintains that a responsible agent's will needs either to have a structure of self-causation or when it does not, it needs to be the downstream product of a self-caused will.²⁶ Though incompatibilists have not focused on this issue, we can reasonably expect that were they to, the debate would parallel the disputes among compatibilists. Historicists will object that when responsible agency is not treated as historical, it looks possible to dramatically manipulate the agent by, for example, controlling which considerations and not others occur to the agent. Semi-structuralists will reply that though (intentional, human) covert manipulation is certainly objectionable, the manipulated agent is still responsible so long as the relevant structural faculties retain their requisite powers. And, of course, historicists will remain unconvinced by this reply.

For my part, I am unsure about the connection between historical intuitions and commitments to incompatibilism. It may well turn out to be the case that compatibilists who endorse a historical account of responsible agency are simply unaware of the work that suppressed incompatibilist intuitions are doing in their thinking about responsibility. But, we would need compelling evidence to ascribe this degree of confusion or even self-deception to so many philosophers. Reliable empirical work could be helpful if it provided some tools for determining the extent to which historical intuitions are indeed widespread, and the extent to which they are connected to with various possible metaphysical commitments relevant to concerns about moral agency. In its absence, it seems to me that we should adopt the picture I suggest above, of treating the history and incompatibilism intuitions as at least *prima facie* distinct.

3.

Let us now consider the issue from a moderate revisionist perspective. Applying a moderate revisionist theory to the agency and history problem yields some immediate methodological advantages. First, a revisionist theory need not deny historicist intuitions, though it also need not hold that we have them. This makes it compatible with whatever empirical results come in about the reactions people have to manipulation cases. Second, even if we do have historicist intuitions, a revisionist theory need not attempt to vindicate them, though it will do so if these intuitions turn out to be justifiable and/or metaphysically innocuous.

We are now in a position to offer a diagnosis of why there is a stalemate at all. The revisionist diagnosis of what has made the debate seem so intractable is simple: lack of clarity about our intuitions generated the problem, and it was exacerbated by a tension between two projects in the theory of moral responsibility— one the one hand, accurately representing or accounting for our commonsense beliefs about freedom and responsibility and on the other, providing a normatively

adequate foundation for our responsibility-characteristic practices, attitudes, and beliefs. Moderate revisionism gives us a way out by abandoning the project of comprehensive intuition vindication.²⁷ While giving an accurate account of folk concepts can be methodologically important, on a moderate revisionist account there is little reason to presume that our folk concepts capture what is normatively adequate and naturalistically plausible. Thus, if historical (or even structural) considerations are to have any force in a satisfactory account of moral responsibility, they must earn it.

The above diagnosis is coupled with a prescription: We should be semi-structuralists. Given some minimal assumptions about the value of responsible agency, there is, thus far, little reason to accept more history than is permitted in a semi-structural theory. What anchors our concern for certain kinds of history, but not others, is an interest in fostering responsible agency.

* * * *

Consider why we ought to care about responsible agency, beyond preserving our intuitions. We might care about responsible agency because it is intimately tied to a respect for the rational valuing capacity of humans. Or, we might care about responsible agency because we want to live in a world where at least other people behave morally. On any credible story— whether Kantian, consequentialist, or other— part of what we care about is that the agents around us have certain kinds of capacities, which include capacities to reason well and to be responsive to a wide range of considerations. For example, agents insensitive to moral reasons (whether at all or in merely in most ordinary contexts) are not susceptible to many of the considerations that move us, and shared discourse with these agents in morally relevant contexts is very difficult. Moreover, agents who altogether lack the capacity to be sensitive to moral considerations cannot be expected to guide their actions in accordance with justified moral norms. Without certain capacities, we can get no traction on morality and morality gets no traction on us. So, inasmuch as there are regular contexts in which

we justifiably want agents to be sensitive to the pull of moral considerations, we have an interest in preserving and encouraging the presence of moral capacities in agents.

At least minimal rationality, sensitivity to justified moral norms, responsiveness to moral reasons, and the presence and normal operation of basic psychological features, including beliefs, pro-attitudes, and intentions, are surely some of the features of agency we are justified in fostering. Minimally, this collection of agential features together provide the kind of structure we have an interest in preserving and encouraging in contexts where what is at stake is the appropriateness of responsibility characteristic practices, attitudes, and beliefs. Call this collection of capacities we are interested in fostering *the basic agential structure of responsibility* (BASR).

Given a justified interest in fostering BASR, it is no feat to see why we ought to care about the history of the agents in cases where it is absent. Recall tracing cases, of the sort exemplified by typical drunk driving examples. In the ordinary case of drunk driving, we can suppose BASR is absent. Nonetheless, there is good reason to hold the agent responsible; at one point he had BASR, and in conditions where the undermining was a reasonably foreseeable result, he either took steps to undermine it or knowingly allowed it to become undermined. What this case shows is that the history of an agent matters at least in cases where the relevant structure is absent, because of our interest in fostering the presence of BASR.

This is not meant to be a point about the preservation of our intuitions but rather about what is normatively adequate. We have a good reason to care about an agent's history in cases where BASR is absent because the history determines whether the agent is an appropriate candidate for the system of practices that tend to encourage the preservation of BASR. So, when an otherwise normal and well-functioning agent has engaged in drunken driving, we have good reason to deploy some of our responsibility-characteristic attitudes (depending on the features of the case: blame, moral indignation, moral outrage). What, precisely, that reason will consist in will vary depending on the

details of the more global account of responsibility that one favors. I am inclined to construe the justification in terms a system of imperfect but generally effective practices that, over time, influence targeted agents to better preserve those capacities that make him or her sensitive to the right considerations.²⁸ If the agent has always been or always will be altogether insensitive to the moral influence embodied in the social interaction of expressions of resent, indignation, outrage (and to the corresponding moral practices), then agent-specific reasons for manifesting the responsibility-characteristic practices and attitudes will become attenuated or eliminated.²⁹ Inasmuch as we are engaged with agents that are or will be sensitive to the justified considerations to which our responsibility-characteristic attitudes, practices, and beliefs direct us, there is a normatively adequate justification for caring about the history of agents lacking BASR. On this account, then, our partial concern for history is anchored in two things; (1) a justified concern in fostering the presence of BASR, and (2) an account of what role the responsibility characteristic practices and attitudes play with respect to our concern for fostering BASR. The latter can presumably be given other construals, though the challenge will be to do so in a way that does not presume the problematic folk concept(s) we are revising.

Of course, for agents that are highly unlikely to have ever had BASR, we may have little justification (relative to an interest in moral responsibility) for having a systematic interest in their history. This is why we do not (and have little reason to) regularly check the history of neighborhood raccoons when we find that they have been involved in some misdeed with the trash; we typically assume that creatures of that sort lack BASR, and are thus inappropriate candidates for the responsibility-characteristic attitudes and practices. However, for creatures we typically assume to possess BASR, we do (and ought to) have a more systematic concern for their histories when we find it missing.

Now consider manipulation cases. Take the case where the agent is manipulated in a way that preserves or grants the agent BASR, where this manipulation is not in the agent's control. Though we may instinctively rebel at treating this as a case of responsible agency, it is difficult to articulate a normatively adequate reason for thinking that the agent ought to be treated as nonresponsible. After all, the agent has all of the capacities we are interested in fostering when we manifest those attitudes and practices characteristic of responsibility. But the fact that something else brought about BASR in the manipulated agent does not mean that the agent is not fully responsible. As Frankfurt pointed out, an agent can be fully responsible without being solely responsible.³⁰ To be sure, if the manipulated agent's actions would not have happened in the absence of manipulation, it seems entirely appropriate to think that at least some of the responsibility for the manipulated agent's action rests on the shoulders of the manipulator. Certainly, we would hold responsible any manipulator in possession of BASR, and we would very likely be justified in doing so. And, in general, we might have good reasons for taking steps to prevent manipulation by manipulators lacking BASR, even if sometimes those manipulations bring about or preserve BASR. However, unless there is some reason to suppose that responsibility must be treated as a zero-sum thing— and I know of no such reason— there is responsibility enough to go around.

Now consider the case where the manipulation robs the agent of BASR. If the agent could³¹ have reasonably foreseen the loss of her basic agential structure of responsibility and avoided it in some suitable sense, the agent is responsible. In cases where the manipulation was not reasonably foreseeable and avoidable, then the agent is not responsible.

In those rarer cases where the manipulation was foreseeable and avoidable, then the agent is responsible. For example, suppose that Enrique, who is otherwise adequately responsive to moral reasons, knows that whenever he is around Carmen, her seductive ways inflame his passion for her so much that he becomes unresponsive to basic moral considerations. In this kind of case, (and

assuming sensitivity to basic moral considerations is part of the basic agential structure) even if Enrique gets manipulated into losing his basic agential structure of responsibility and goes on to kill one of Carmen's other suitors, Enrique is still responsible because he could have reasonably foreseen and avoided the loss his basic agential structure of responsibility.³²

Now consider "covert non-constraining control" cases, or so-called Brave New World examples. As in Huxley's novel, these agents may be raised or engineered to prefer their available choices.³³ Such agents are systematically and subtly manipulated such that BASR is intact, but that at least intuitively, the agents are not responsible. Incompatibilists often emphasize the difficulties compatibilists have in either denying the apparently widespread intuition that, in Brave New World cases, agents are not responsible, or (if they do accept the intuition that they are not responsible), in accounting for the relevant difference between Brave New World cases and determinism.³⁴ Given standard forms of compatibilism, neither horn of the dilemma seems attractive. On one horn, it is difficult to dismiss Brave New World cases because too many people seem willing to say they are not cases of responsibility. On the other, it is a not insignificant challenge to explain the difference between determinism (or even ordinary causal interactions with the world) and the systematic, covert manipulation that takes place in Brave New World-style examples.³⁵

Here, the moderate revisionist has an answer that splits the difference. The intuition (inasmuch as we have it) that Brave New World agents are not responsible is likely an artifact of historical impulses in ordinary thinking about responsibility. However, once we get clear on what is normatively adequate in our collection of historicist impulses, and that a history of manipulation does not, by itself, make you an inappropriate target of the responsibility characteristic practices and attitudes, the relevant difference for responsibility is the presence or absence of BASR. If BASR is present in a Brave New World case, the agent ought to be counted as a responsible agent because she has the capacities we are justified in fostering through moral influence. Of course, where BASR

is not present, the standard interest in history comes into play. So, *Brave New World* cases fail to show that there is no difference between determinism and the kind of agency that can provide the normative basis for responsibility-characteristic practices, attitudes, and beliefs. Rather, what they show is the problem with the pervasive assumption that our ordinary intuitions about responsibility do a good job of tracking what is, in fact, normatively justifiable.

To briefly summarize: Fischer and Ravizza have challenged semi-structuralists to give a principled reason to care for some histories and not others. Our concern for BASR does that. Though there may be strong intuitions that support a historical theory, until we have an argument for why those intuitions track a justified concern, it would be sheer prejudice to care about more history than is offered by semi-structuralism.

This proposal maintains that many of our responsibility-characteristic practices and attitudes can be justified in virtue of an interest in fostering responsible agency. That many of our practices and attitudes can be justified in light of the goal of moral influence does not mean that there are not further justifications available, justifications that may overlap, add wrinkles to, or constrain the moral influence justification offered here. But, such additions will not change the basic fact that moral influence is justified and adequate grounds for a semi-structural theory. Certainly, the precise way we understand the moral influence theory will depend on the details of the broader normative theory in which it is embedded. I have already emphasized that the account I offer can be cast in both Kantian and Consequentialist lights, though presumably other characterizations are possible as well. Depending on one's broader normative commitments, the justificatory substrate I have offered will take on a different hue and will be developed in a different way. However, the fact that many of our responsibility-characteristic practices can be justified because of their role in moral influence provides a secure foundation for many of our responsibility-characteristic practices and attitudes, even if there are other justification strategies that can augment this account.³⁶

For example, consider an independent account of the wrongness of killing. Suppose that what makes killing wrong, when it is wrong, is that it robs that agent of a valuable future, where that is understood to mean a future in which the agent values something— itself, others, activities, or things.³⁷ On this view, it would not be wrong to kill someone if they were incapable of having a valuable future in this sense. On the other hand, under normal circumstances it would be wrong to kill a suicidal woman even if she did not value anything, because if she were to continue living she would (very likely) come to value life, herself, others, or some other thing. If we marry the account of the wrongness of killing with the moderate revisionist theory, we get an account that can explain when an agent is an apt target for the praise and blame and what makes the action wrong (and thus blameworthy). This example shows that accepting the moderate revisionist proposal need not commit us to accounting for the wrongness of acts solely in terms of forward-looking considerations. Since the justification of praise and blame I have offered can be independent of the account of the wrongness of killing (though it may not be for a strict consequentialist implementation of the moral influence theory), there are likely to be interesting divergences between when someone has done something wrong and when the agent is accountable.

4.

I have offered a proposal for with three parts: (1) a thin account of responsible agency (BASR), (2) a minimal account of the justification of a large subset of our responsibility practices (moral influence), and (3) a moderate revisionist approach in which (1) and (2) are embedded. Given that the revisionism is stipulated, the most natural targets of criticism are (1) and (2): the accounts of responsible agency and moral influence. I will take up some criticisms of the moral influence part before concluding with reflections on a possible objection directed at the account of responsible agency.

Two objections to the moral influence part of the proposal come from opposite directions: the proposed account might, depending on the critic, seem trivially true or obviously false. A critic might object that the account is largely toothless, because few would dispute that moral influence accounts can secure as much as I have suggested.³⁸ Minimally, this criticism would require engagement with the complex semantic issues to which I mention in section 1 of this paper, for part of what is at stake is whether a revisionist moral influence account of responsible agency gets us what we want from a theory of responsible agency. In turn, this issue depends in part on what we mean by “responsible agency.” Until such critical engagement is offered, it is not so obvious that this account would be so readily accepted by the various partisans of the free will debates. Moreover, a wide range of philosophers have argued that if incompatibilism and the thesis of determinism were both true, we would be faced with the collapse of praise and blame³⁹, desert⁴⁰, and/or moral responsibility.⁴¹ It would be difficult to hold this range of views and consistently maintain that the account of responsible agency I have offered — one immune to determinism — is also true. Indeed, it is indicative that Fischer and Ravizza reject incompatibilism for something like this reason: they hold that it is implausible to believe, as they think incompatibilists must, that we would have to abandon our reactive attitudes and associated practices if we were to learn that determinism is true.⁴² Hence, though the account is (I hope) largely true, it is certainly not trivially so.

It is notable that when philosophers have discussed moral influence accounts, they have largely been critical.⁴³ This brings us to the reverse criticism, that though moral influence accounts have teeth, those teeth are in the mouth of a dead horse. There is an influential group of criticisms that might rightly be thought to have finished off moral influence accounts of responsibility, and so one might object to the proposed account on the grounds that moral influence is not a credible basis for a theory of responsibility.⁴⁴ Elsewhere, I have argued that these criticisms can be overcome, and that a moral influence theory has the resources to justify a wide range of responsibility-characteristic

attitudes and practices.⁴⁵ The key idea is that a satisfactory moral influence theory need not hold that every deployment of the responsibility-characteristic attitudes and practices promotes some desirable outcome. Rather, it simply requires that our messy system of responsibility-characteristic practices and attitudes generally, and over time, achieves our goals better than alternatives. What we need here, though, is even more minimal. The proposal we have been considering merely requires that the capacities presupposed by our system of moral influence are valuable enough to justify concern for them. Even if a general moral influence account of responsibility is unsatisfactory, as long as our concern for BASR is justified, we have all we need for the proposed account of responsible agency. In short, a moderate revisionist account of responsible agency is neither toothless nor obviously dead.

A different sort of criticism is likely to be pushed by historicists. Historicists might concede that nothing in BASR *as I have outlined it* contains historicist elements, but that a complete and true account of it would have elements with ineradicably historical commitments.⁴⁶ Something similar might be said about the moral influence justification of our responsibility practices and attitudes. Given that the proposal does not rule out other possible, overlapping justifications, a further account that justifies our responsibility-characteristic practices might rely on some deeply historical element of agency. This possibility cannot be ruled out. But, this move would signal an important shift in the historicist project. Historicists— i.e., the majority view— would no longer be able to simply rely on the apparent intuitiveness of historical considerations. They would now have the burden of showing just what the alleged historical element is and why it plays a justified (or minimally, necessary) role in our evaluation of agents. And, if they did, then revisionists should change their views accordingly— they are, after all, revisionists.

However, if the most historicists can show is that there is a historical element to our thinking without showing that this element is justified or necessary in our evaluation of agents, this will not

be enough. Non-historicists will object that the historical element is gratuitous and unacceptably prejudicial: gratuitous because the element plays no role in justifying our concern with responsible agency, and prejudicial because (at least potentially) blame and punishment could come to be meted out on the basis of nothing more than a contingent accident of our culturally-inherited way of thinking about responsibility. In contrast, the account I have offered can explain what role history should play in our evaluation of agents and why it is justified. This is something no historical account has thus far done.

For moderate revisionists the concern with the possible gratuitousness is an obvious consequence of the standard of normative adequacy. However, it is important to note that this concern is not limited to revisionists. Compatibilists have long pressured libertarians (those who think we have free will and it requires indeterminism) to explain what work the postulated indeterminism does in getting what libertarians are after (whether free will, moral responsibility, dignity, objective value, etc.).⁴⁷ By and large, incompatibilists seem to have accepted this constraint.⁴⁸ It would therefore be puzzling to apply this standard to accounts of agency in which the necessity of indeterminism is at stake, but not to accounts where the necessity of history is at stake. Inasmuch as the concern has widespread currency in debates about free will and moral responsibility, the argument I have made here applies to more than just revisionist accounts.

* * * *

A key part of the modern philosophical imagination about free will and moral responsibility is the idea that the integrity of our responsibility-characteristic practices and attitudes depends on the vindication of our self-conception as free and responsible agents in the world. The debate between traditional compatibilists and incompatibilists has been over the characterization of that self-conception. The difficulty with trying to shoehorn our best philosophical theories into the constraints of our presumably contingent and culturally inherited intuitions is well-illustrated by the

intuition stand-off that characterizes the history/structure debate: too often our intuitions clash with the normative aims of our theories. The alternative I propose is one that admits of moderate revisions to our self-conception, while holding that we can justify much of what we were concerned to account for in the first place. Thus, we can be open to the idea that ordinary thinking about responsibility may well be historical in the way historicists insist, while maintaining that an adequate theory will depart from the folk concept in counterintuitive ways. Counterintuitiveness is thus a predictable consequence of the approach. That is, we will find the theory counterintuitive in just those instances where the theory revises away from the elements of our practices, attitudes, and beliefs that are either unjustifiable or naturalistically implausible. Of course, the point and the promise of a moderately revisionist approach is that it captures what matters, and no more. That should be enough.⁴⁹

¹ Several points of clarification may be helpful. First, I say “incompatible in some deep way” to recognize the possibility that one might hold that some superficial aspects or notions of responsibility are incompatible with determinism, without a core or central notion being incompatible. Second, I use ‘incompatibilism’ and ‘compatibilism’ to refer to the views about the relationship between moral responsibility and determinism. These terms are also frequently used to refer to views about free will and determinism. I treat free will as a condition of morally responsible agency (the “freedom condition,” as I will also call it). Theories that take contrasting stands on free will/determinism and moral responsibility/determinism (e.g., “semi-compatibilism”) will be understood in terms of the analysis of responsibility. This gives my discussion a particular slant, but it is not meant as a substantive claim that restricts concerns about free will to concerns about moral responsibility.

² For various defenses of this claim, see Peter Van Inwagen, *An Essay on Free Will* (New York: Oxford University Press, 1983), as well as Peter Van Inwagen, “Free Will Remains a Mystery,” in *Philosophical Perspectives*, ed. James Tomberlin (Boston: Blackwell, 2000), and Ted Warfield, “Causal Determinism and Human Freedom Are Incompatible: A New Argument for Incompatibilism,” in *Philosophical Perspectives*, ed. James Tomberlin (Boston: Blackwell, 2000). See also Laura Waddell Ekstrom, *Free Will: A Philosophical Study* (Boulder, Colorado: Westview Press, 2000), Robert Kane, *The Significance of Free Will* (Oxford: Oxford, 1996), Timothy O’Connor, *Persons and Causes* (New York: Oxford, 2000), Derk Pereboom, *Living without Free Will* (Cambridge: Cambridge, 2001), and Galen Strawson, “The Impossibility of Moral Responsibility,” *Philosophical Studies* 75 (1994). For doubts about the metaphysical and scientific plausibility of one or another form of libertarianism, J.J.C. Smart, “Free Will, Praise, and Blame,” *Mind* 70 (1961), Saul Smilansky, *Free Will and Illusion* (New York: Clarendon, 2000), Strawson, “The Impossibility of Moral Responsibility,” Van Inwagen, “Free Will Remains a Mystery,” Henrik Walter, *Neurophilosophy of Free Will: From Libertarian Illusions to a Concept of Natural Autonomy* (Cambridge, Mass.: MIT Press, 2001), Patricia Churchland, *Brainwise* (Cambridge, Mass.: MIT, 2002), 201-237.

³ For discussion and defense of these claims, see Manuel Vargas, “Responsibility and the Aims of Theory: Strawson and Revisionism,” *Pacific Philosophical Quarterly* 85, no. 2 (2004), Manuel Vargas, “The Revisionist’s Guide to Responsibility,” *Philosophical Studies* (forthcoming), and Manuel Vargas, “Moral Influence, Moral Responsibility,” (Under review.)

⁴ Various forms of compatibilist conditional analyses of ‘can’ have had this form. For more on this and alternative forms of revisionism, see “Vargas, “The Revisionist’s Guide to Responsibility,” and Manuel Vargas, “Compatibilism Evolves? On Some Varieties of Dennett Worth Wanting,” *Metaphilosophy* (forthcoming).

⁵ In adopting the normative standard I am assuming that realism about moral properties is compatible with naturalism. Even if one rejects this assumption there are still two options. First, one might accept the standard of naturalistic plausibility, but take a different view about how to talk about the issue of a revisionist picture of moral responsibility. Second, one might pursue a kind of revisionism compatible with only the standard of normative adequacy. Either way, these would count as revisionist theories, though not of the sort I pursue.

⁶ If one has a view about the semantics of moral properties where it turns out that (1) there really is a property of responsibility, (2) it is not ontologically troublesome, but (3), our thinking about that property is error-ridden, then one might believe that we should revise our thinking so that it properly tracks the features of the property of responsibility. But notice, even if you think there is a property of responsibility, that property will have to satisfy the two standards I go on to adopt, or else it is hard to see why we should have any investment in a theory of that property (i.e., the property of responsibility, strictly speaking). See Vargas, "The Revisionist's Guide to Responsibility,"

⁷ John Martin Fischer and Mark Ravizza, *Responsibility and Control: A Theory of Moral Responsibility* (New York: Cambridge University Press, 1998), p. 207.

⁸ For Mele, the relevant notion of compulsion is one where the acquisition of the relevant features of the agent (S's pro-attitudes) is brought about in a way that "bypasses S's capacities for control over his mental life; and the bypassing issues in S's being practically unable to shed [the relevant pro-attitude(s)]; and the bypassing was not itself arranged (or performed) by S; and S neither presently possesses nor earlier possessed pro-attitudes that would supply his identifying with [the relevant pro-attitude(s)], with the exception of the pro-attitudes that are themselves practically unsheddable products of unsolicited by-passing" (172). See Alfred Mele, *Autonomous Agents: From Self-Control to Autonomy* (New York: Oxford University Press, 1995).

⁹ Michael Bratman has noted that Fischer and Ravizza need to say more about how understanding the idea of "taking responsibility" in this way makes reference to a mechanism (See Michael Bratman, "Fischer and Ravizza on Moral Responsibility and History," *Philosophy and Phenomenological Research* LXI, no. 2 (2000), esp. 454). One set of suggestions can be found in John Martin Fischer, "Responsibility, History, and Manipulation," *Journal of Ethics* 4, no. 4 (2000). There, Fischer writes that "we take responsibility in the first instance for *kinds* of mechanisms or *kinds* of processes that issue in our behavior" (p. 390), and he claims that it is a condition of taking responsibility for these kinds of mechanisms that the agent view her actions in the world as resulting from his or her "own deliberations, choices, or bodily movement."

Together, I take these things to suggest that the idea is that the agent must think of his actual dispositions, habits, and

reasoning mechanisms that lead to his various deliberations, choices, and actions *as his own*. In cases where manipulation has occurred, Fischer has claimed in John Martin Fischer, "Précis of Responsibility and Control: A Theory of Moral Responsibility," *Philosophy and Phenomenological Research* LXI, no. 2 (2000), that the agent is not typically taking responsibility for the right kind of mechanism (that is, a manipulated mechanism)— unless, of course, the agent is told of the manipulation. Note, though, that depending on how ownership is understood, this by itself may not be sufficient to ground the claim that responsibility is essentially historical. That is, if ownership can be accounted for by (for example) simply having a certain structure of agency, none of the preceding is yet sufficient to warrant the claim that responsibility is "essentially historical." That is why the business about *kinds* of mechanisms seems to matter. If the kind of mechanism has to do with its causal history, then it looks like history can get in this way. But in another place (John Martin Fischer, "Replies," *Philosophy and Phenomenological Research* LXI, no. 2 (2000)) Fischer and Ravizza seem to back off of a similar strong degree of knowledge when they claim that "we have in mind a way of individuating mechanisms which does not require that the agent know everything about the causal origins of its inputs or its inner working" (p. 476). See also David Zimmerman, "Reasons-Responsiveness and Ownership-of-Agency," *Journal of Ethics* 6 (2002).

¹⁰ Harry Frankfurt, "Freedom of the Will and the Concept of a Person," *Journal of Philosophy* 68, no. 1 (1971) reprinted in Harry G. Frankfurt, *The Importance of What We Care About: Philosophical Essays* (Cambridge: Cambridge University Press, 1988). Gerald Dworkin's work on autonomy (see Gerald Dworkin, *The Theory and Practice of Autonomy* (New York: Cambridge, 1988)) is similar in this respect; though Mele highlights some complexities in interpreting Dworkin's account as ahistorical. See Mele, *Autonomous Agents: From Self-Control to Autonomy* p. 148.

¹¹ Though it makes a useful illustration of structural theories, Frankfurt's account of addiction is controversial. For a critical survey of recent work on addiction, see Gideon Yaffe, "Recent Work on Addiction and Responsible Agency," *Philosophy and Public Affairs* 30, no. 2 (2002).

¹² Zimmerman labels these "process-historical" and "source-historical." The former refers to the temporal extendedness of certain structural features and the latter refers to what I have in mind by historical theories. See Zimmerman, "Reasons-Responsiveness and Ownership-of-Agency," esp. remarks on 210 and n.34. What Zimmerman and I call structural and historical theories, Fischer and Ravizza in *Responsibility and Control* (New York: Cambridge, 1998) Fischer and Ravizza, *Responsibility and Control: A Theory of Moral Responsibility* call "mesh theories" and "historical theories" (183-4) and Mele in *Autonomous Agents*, calls "internal and external conceptions of psychological autonomy" (146-149).

¹³ What I am calling a “semi-structural” account was suggested but not endorsed by Bratman, “Fischer and Ravizza on Moral Responsibility and History,” Ishtiyaque Haji has also suggested (also without explicitly endorsing) the idea that the Frankfurtian might move to something that seems to be a relative to the semi-structuralist theory I have just described. See Ishtiyaque Haji, “On Responsibility, History, and Taking Responsibility,” *Journal of Ethics* 4, no. 4 (2000). Haji suggests that we can “distinguish between concerns of responsibility-grounding control and those of ultimacy, roughly, those factors in virtue of which an agent's springs of action are truly an agent's own” (399). The idea seems to be that a Frankfurtian might admit that history matters for questions about an action being “truly an agent's own,” but that these issues are independent of the assessment of responsibility, which can be accounted for in Frankfurtian terms. As will become evident, I am less optimistic about a purely structural account of responsibility though I am friendly to the idea that issues of ultimacy may well pull apart from issues of responsibility.

¹⁴ There are several further responses the structuralist might pursue. For instance, the structuralist could argue that we ought to hold the considered agent responsible only for being drunk, but not responsible for any of the consequences that follow from the drunkenness. Though clearly a departure from ordinary moral thinking, this approach would carry the benefit of avoiding some of the problems that go under the heading of “moral luck” (e.g., that we would treat worse a drunk driver who got “unlucky” and killed someone while drunk driving as opposed to another drunk driver who got lucky and didn’t kill anyone). It seems to me, however, that there are good reason to hold drunk drivers responsible for more than their drunkenness. At the very least, typical agents who end up driving drunk at some point knew (or culpably didn’t know) that they were creating conditions under which horrible results have a considerably higher likelihood of happening. Consequentialist and Kantian explanations may augment this idea by emphasizing, in the consequentialist analysis, the beneficial consequences of holding agents responsible for the consequences of their drunken driving, or in the latter case, the failure of the drunken driver to show suitable respect for oneself and others by failing to recognize the hazard one has become. Of course, more would need to be said to address the further issue of moral luck.

¹⁵ On an alternative example emphasizing reasons-responsiveness, we might suppose that the nefarious neurosurgeon gives Octavio a strong but not an overriding or irresistible desire (leaving his reasons-responsiveness intact).

¹⁶ See Fischer, “Replies,” and Fischer, “Responsibility, History, and Manipulation,” It is notable that historicist “hard incompatibilists” also rely on manipulation cases. See, for example, G. Strawson Strawson, “The Impossibility of Moral Responsibility,” and Pereboom, *Living without Free Will*

¹⁷ See Bratman, “Fischer and Ravizza on Moral Responsibility and History,”

¹⁸ Might a historicist agree? On Mele's account Octavio's acquisition of the desire to whistle will not undermine his responsibility as long as the desire is "sheddable" (see note 8). So, even if Octavio's desire was acquired by manipulation, acquisition of sheddable desires does not amount to compulsion. This gets at an interesting issue about whether Mele's account is historical in the sense I mean. If Mele thinks that the paradigmatic or model agential structure required for responsibility is one in which an agent's relevant desires are sheddable, then what I have been calling his "negative historical condition" is really the tracing condition of a semi-structural account. On the other hand, if Mele does not think that there is any such requirement on the basic agential structure of responsibility, then his theory does count as a genuinely historical one. A case that is worth thinking about in this context is one where an agent has only sheddable desires— if Mele thinks that these structural properties (plus whatever other structural properties he thinks are required for responsible agency) are enough to make the agent a responsible agent, then on my way of thinking about these things, he is a semi-structuralist. Given his stated account of compulsion, I am inclined to think that this is what he should say. However, if Mele thinks that there is some further condition that really does require that responsible agents *always* have some historical property, I suspect we could cook up a case that would do the work that Octavio does against, for example, Fischer and Ravizza.

¹⁹ Bratman, "Fischer and Ravizza on Moral Responsibility and History,"

²⁰ Fischer, "Replies," p. 473.

²¹ A structuralist could accept that our treatment of history should be symmetrical and then go on to argue that it should never matter. As we have already seen, though, a structuralist will have difficulties with ordinary tracing cases. See n.14 for more, though.

²² A similar point is emphasized in Zimmerman, "Reasons-Responsiveness and Ownership-of-Agency," p. 215.

²³ For example, Fischer has thought that "one reason why Harry Frankfurt has so frequently reiterated the view that responsibility is not a matter of history is precisely because he is afraid of falling into the clutches of incompatibilism." See John Martin Fischer, "Chicken Soup for the Semi-Compatibilist Soul: Reply to Haji and Kane," *Journal of Ethics* 4, no. 4 (2000), esp. 407. See also John Martin Fischer and Mark Ravizza, "Responsibility and History," in *Midwest Studies in Philosophy*, ed. Peter A. French, Theodore E. Jr. Uehling, and Howard K. Wettstein (Notre Dame: Notre Dame, 1994), esp. p. 444. Richard Double, who is not a compatibilist, has made similar remarks. See Richard Double, *The Non-Reality of Free Will* (New York: Oxford University Press, 1991), pp. 56-7.

²⁴ For evidence of this point, see Zimmerman, “Reasons-Responsiveness and Ownership-of-Agency,” esp. n. 2 and n. 13.)

²⁵ Ekstrom, *Free Will: A Philosophical Study* pp.245-7.

²⁶ Strawson, “The Impossibility of Moral Responsibility,”

²⁷ In this respect, the project shares something with David Lewis’ remarks against “perfectionism” in David Lewis, “Humean Supervenience Debugged,” *Mind* 103 (1994) reprinted in David Lewis, *Papers in Metaphysics and Epistemology* (Cambridge, U.K. ; New York: Cambridge University Press, 1999).

²⁸ I defend this view in greater detail in Vargas, “Moral Influence, Moral Responsibility,”

²⁹ There is a fascinating literature, relevant to some of these issues, that has grown out of P. F. Strawson, “Freedom and Resentment,” *Proceedings of the British Academy* XLVIII (1962). In particular, see Lawrence Stern, “Freedom, Blame, and the Moral Community,” *Journal of Philosophy* 71, no. 3 (1974), Gary Watson, “Responsibility and the Limits of Evil,” in *Responsibility, Character, and the Emotions*, ed. Ferdinand David Schoeman (New York: Cambridge, 1987), R. Jay Wallace, *Responsibility and the Moral Sentiments* (Cambridge, Mass.: Harvard University Press, 1994), and Michael McKenna, “The Limits of Evil and the Role of Moral Address,” *Journal of Ethics* 2, no. 2 (1998). I discuss revisionist construals of the Strawsonian project in “Responsibility and the Aims of Theory.”

³⁰ Frankfurt, *The Importance of What We Care About* p. 25 n.10.

³¹ Revisionism about responsibility will likely entail revisionism about a network of responsibility-relevant concepts. In particular, the responsibility-relevant notions of ‘can’ and ‘avoidability’ may need revision as well, depending on how their current semantics are properly understood. However, it is worth noting that a number of philosophers with broadly incompatibilist leanings have acknowledged that there are perfectly intelligible senses of ‘can’ and other “power” terms that are not threatened by determinism. See, for example, Kane, *The Significance of Free Will*, Honderich *A Theory of Determinism*, and Smilansky, *Free Will and Illusion*. Thus, even if the responsibility-relevant notion of these words is *not* immune to the threat of determinism, there is no reason why we cannot revise them so that they are immune to these difficulties.

³² Mele discusses similar cases in *Autonomous Agents*, Ch. 9.

³³ Aldous Huxley, *Brave New World* (New York: Perennial, 1942).

³⁴ See Robert Kane, “Non-Constraining Control and the Threat of Social Conditioning,” *Journal of Ethics* 4, no. 4 (2000), Pereboom *Living Without Free Will*, and Zimmerman “Reasons-Responsiveness and Ownership-of-Agency.” They agree

about the historicist pull of these kinds of cases, though they disagree about the prospects for libertarianism and compatibilism.

³⁵ This is a line of argument that has received particular emphasis in the work of Derk Pereboom, in the form of his “Generalization Argument.” See *Living Without Free Will*, pp. 110-120. Whether it and argument like it can satisfactorily be answered is contentious. And of course, a number of philosophers (including some I have been discussing) have offered accounts that are intended to show that there is a principled difference between the effects of determinism and systematic covert manipulation. See, for example, Mele’s *Autonomous Agents*, chapters 9-10, and Fischer and Ravizza’s *Responsibility and Control*, especially chapter 8. As my own account should make clear, I too think that there is a way to make a principled distinction between determinism and at least some manipulation cases.

³⁶ Despite all I have said, it could turn out that a moral influence account ends up generating a principled concern for history. For example, suppose that it strikes many people as unfair to blame or punish brainwashed agents who do have the BASR, and suppose that this view did not depend on any factual errors or confusions. It could turn out that the perceived unfairness of blaming or punishing in instances such as these could have a corrosive effect on the stability and efficacy of the entire system of moral influence. If so, (all other things being equal), this might entail that that we adopt a historicist picture of responsible agency. However, I am inclined to think that counterintuitive results in cases as uncommon as the sorts of brain washing we have been considering are unlikely to have so strong an effect on the stability and efficacy of a plausible system of moral influence— but this is an *a posteriori* matter about which I may well be wrong. My thanks to a referee for raising this possibility.

³⁷ See Don Marquis, “Why Abortion Is Immoral,” *Journal of Philosophy* 86, no. 4 (1989) for a defense of a view similar to the one I am proposing. It seems to me, however, that Marquis’ account is lacking an explanation of what sorts of entities can and cannot have a valuable future.

³⁸ Thanks to Derk Pereboom for suggesting this criticism.

³⁹ Smart, “Free Will, Praise, and Blame,”

⁴⁰ G. Strawson “The Impossibility of Moral Responsibility” and Saul Smilansky *Free Will and Illusion*.

⁴¹ Fischer and Ravizza, *Responsibility and Control*, van Inwagen *An Essay on Free Will*, Pereboom, *Living Without Free Will*, G. Strawson “The Impossibility of Moral Responsibility” and Saul Smilansky *Free Will and Illusion*.

⁴² *Responsibility and Control*.

⁴³ Daniel Dennett, *Elbow Room* (Cambridge: MIT, 1984) appears to be the lone exception— and a widely criticized one at that. For example, see Gary Watson, “Review of *Elbow Room*,” *The Journal of Philosophy* 83, no. 9 (1986) and Gerald Dworkin, “Review of *Elbow Room*,” *Ethics* 96, no. 2 (1986).

⁴⁴ Some examples include Strawson “Freedom and Resentment,” Jonathan Bennett, “Accountability,” in *Philosophical Subjects*, ed. Zak Van Straaten (New York: Clarendon, 1980); and T. M. Scanlon, “The Significance of Choice,” in *Tanner Lectures* (1988).

⁴⁵ Vargas “Moral Influence, Moral Responsibility”.

⁴⁶ Thanks to Al Mele for suggesting this line of response.

⁴⁷ See Gary Watson, “Free Action and Free Will,” *Mind* 96 (1987), Michael Bratman, “Nozick on Free Will,” ed. David Schmidtz (New York: Cambridge, 2002), and Alfred Mele, “Kane, Luck, and the Significance of Free Will,” *Philosophical Explorations* 2 (1999), though this list is by no means exhaustive. It is an interesting question whether compatibilists can hold on to their non-revisionism while pushing the non-gratuity point. I am doubtful, which is why I think many compatibilists are latent revisionists— though see “The Revisionist’s Guide to Responsibility,” “Responsibility and the Aims of Theory,” and “Compatibilism Evolves?” for some of the complexities involved.

⁴⁸ For example, see Kane *The Significance of Free Will*, Robert Kane, “Responsibility, Luck, and Chance: Reflections on Free Will and Indeterminism,” *Journal of Philosophy* XCVI, no. 5 (1999), and Randolph Clarke, *Libertarian Accounts of Free Will* (Oxford: Oxford, 2003).

⁴⁹ My thanks for valuable comments and criticisms goes to Michael Bratman, Christopher Grau, Alan “H-Bomb” Hájek, Agnieszka Jaworska, Janet Levin, Simon May, Michael McKenna, Al Mele, Cei Maslen, Derk Pereboom, Gideon Yaffe, as well as audiences where ancestors of this paper were presented, including the 2004 Pacific Meeting of the American Philosophical Association, Florida State, San José State University, the University of New Mexico, the University of San Francisco, University of Southern California, and the University of California, Irvine. Thanks also to the University of San Francisco Fleishhacker Family Endowment and the California Institute of Technology for financial support while working on parts of this paper.

References

- Bennett, Jonathan. "Accountability." In *Philosophical Subjects*, edited by Zak Van Straaten. New York: Clarendon, 1980.
- Bratman, Michael. "Fischer and Ravizza on Moral Responsibility and History." *Philosophy and Phenomenological Research* LXI, no. 2 (2000): 453-58.
- . "Nozick on Free Will." edited by David Schmidtz, 155-74. New York: Cambridge, 2002.
- Churchland, Patricia. *Brainwise*. Cambridge, Mass.: MIT, 2002.
- Clarke, Randolph. *Libertarian Accounts of Free Will*. Oxford: Oxford, 2003.
- Dennett, Daniel. *Elbow Room*. Cambridge: MIT, 1984.
- Double, Richard. *The Non-Reality of Free Will*. New York: Oxford University Press, 1991.
- Dworkin, Gerald. "Review of *Elbow Room*." *Ethics* 96, no. 2 (1986): 423-25.
- . *The Theory and Practice of Autonomy*. New York: Cambridge, 1988.
- Ekstrom, Laura Waddell. *Free Will: A Philosophical Study*. Boulder, Colorado: Westview Press, 2000.
- Fischer, John Martin. "Chicken Soup for the Semi-Compatibilist Soul: Reply to Haji and Kane." *Journal of Ethics* 4, no. 4 (2000): 404-07.
- . "Précis of Responsibility and Control: A Theory of Moral Responsibility." *Philosophy and Phenomenological Research* LXI, no. 2 (2000): 441-46.
- . "Replies." *Philosophy and Phenomenological Research* LXI, no. 2 (2000): 467-80.
- . "Responsibility, History, and Manipulation." *Journal of Ethics* 4, no. 4 (2000): 385-91.
- Fischer, John Martin, and Mark Ravizza. *Responsibility and Control: A Theory of Moral Responsibility*. New York: Cambridge University Press, 1998.

- . “Responsibility and History.” In *Midwest Studies in Philosophy*, edited by Peter A. French, Theodore E. Jr. Uehling and Howard K. Wettstein, 430-51. Notre Dame: Notre Dame, 1994.
- Frankfurt, Harry. “Freedom of the Will and the Concept of a Person.” *Journal of Philosophy* 68, no. 1 (1971): 5-20.
- Frankfurt, Harry G. *The Importance of What We Care About: Philosophical Essays*. Cambridge: Cambridge University Press, 1988.
- Haji, Ishtiyaque. “On Responsibility, History, and Taking Responsibility.” *Journal of Ethics* 4, no. 4 (2000): 351-56.
- Huxley, Aldous. *Brave New World*. New York: Perennial, 1942.
- Kane, Robert. “Non-Constraining Control and the Threat of Social Conditioning.” *Journal of Ethics* 4, no. 4 (2000): 357-60.
- . “Responsibility, Luck, and Chance: Reflections on Free Will and Indeterminism.” *Journal of Philosophy* XCVI, no. 5 (1999): 217-40.
- . *The Significance of Free Will*. Oxford: Oxford, 1996.
- Lewis, David. “Humean Supervenience Debugged.” *Mind* 103 (1994): 473-90.
- . *Papers in Metaphysics and Epistemology*. Cambridge, U.K. ; New York: Cambridge University Press, 1999.
- Marquis, Don. “Why Abortion Is Immoral.” *Journal of Philosophy* 86, no. 4 (1989): 183-202.
- McKenna, Michael. “The Limits of Evil and the Role of Moral Address.” *Journal of Ethics* 2, no. 2 (1998): 123-42.
- Mele, Alfred. *Autonomous Agents: From Self-Control to Autonomy*. New York: Oxford University Press, 1995.
- . “Kane, Luck, and the Significance of Free Will.” *Philosophical Explorations* 2 (1999): 96-104.

- O'Connor, Timothy. *Persons and Causes*. New York: Oxford, 2000.
- Pereboom, Derk. *Living without Free Will*. Cambridge: Cambridge, 2001.
- Scanlon, T. M. "The Significance of Choice." In *Tanner Lectures*, 1988.
- Smart, J.J.C. "Free Will, Praise, and Blame." *Mind* 70 (1961): 291-306.
- Smilansky, Saul. *Free Will and Illusion*. New York: Clarendon, 2000.
- Stern, Lawrence. "Freedom, Blame, and the Moral Community." *Journal of Philosophy* 71, no. 3 (1974): 72-84.
- Strawson, Galen. "The Impossibility of Moral Responsibility." *Philosophical Studies* 75 (1994): 5-24.
- Strawson, P. F. "Freedom and Resentment." *Proceedings of the British Academy* XLVIII (1962): 1-25.
- Van Inwagen, Peter. *An Essay on Free Will*. New York: Oxford University Press, 1983.
- . "Free Will Remains a Mystery." In *Philosophical Perspectives*, edited by James Tomberlin, 1-20. Boston: Blackwell, 2000.
- Vargas, Manuel. "Compatibilism Evolves? On Some Varieties of Dennett Worth Wanting." *Metaphilosophy* (forthcoming).
- . "Moral Influence, Moral Responsibility." Under review.
- . "Responsibility and the Aims of Theory: Strawson and Revisionism." *Pacific Philosophical Quarterly* 85, no. 2 (2004): 218-41.
- . "The Revisionist's Guide to Responsibility." *Philosophical Studies* (forthcoming).
- Wallace, R. Jay. *Responsibility and the Moral Sentiments*. Cambridge, Mass.: Harvard University Press, 1994.
- Walter, Henrik. *Neurophilosophy of Free Will: From Libertarian Illusions to a Concept of Natural Autonomy*. Cambridge, Mass.: MIT Press, 2001.

Warfield, Ted. "Causal Determinism and Human Freedom Are Incompatible: A New Argument for Incompatibilism." In *Philosophical Perspectives*, edited by James Tomberlin, 167-80. Boston: Blackwell, 2000.

Watson, Gary. "Free Action and Free Will." *Mind* 96 (1987): 145-72.

———. "Responsibility and the Limits of Evil." In *Responsibility, Character, and the Emotions*, edited by Ferdinand David Schoeman, 256-86. New York: Cambridge, 1987.

———. "Review of *Elbow Room*." *The Journal of Philosophy* 83, no. 9 (1986): 517-22.

Yaffe, Gideon. "Recent Work on Addiction and Responsible Agency." *Philosophy and Public Affairs* 30, no. 2 (2002): 178-221.

Zimmerman, David. "Reasons-Responsiveness and Ownership-of-Agency." *Journal of Ethics* 6 (2002): 199-234.