How to Manipulate an Incompatibilistically Free Agent

Penultimate version – please cite published version, *APQ* 49(2): 139-49

A prominent criticism of compatibilist theories of moral responsibility is that they do not deal adequately with cases of manipulation, and particularly with induced desires. So, for example, accounts which give the conditions for responsibility in terms of the reasons-responsiveness of the mechanism governing one's decisions, as Fischer and Ravizza (1998) do, seem to leave the door open to cases such as the following.

Suppose Billy's mother is a mad scientist, and concerned that their neighbours think him a good moral citizen. She therefore implants a chip in his brain, which is activated whenever anyone can see him, and which works to force him to see the moral reasons there are and act on them appropriately. Then when the chip is active, Billy will be acting on a reasons-responsive mechanism, and so it seems the compatibilist would have to consider him responsible for his actions.[1] But, given that when the chip is inactive, Billy is a thoroughly disgraceful moral agent, and that the chip's activity causes him to make decisions he otherwise would not, this seems like a clear case of manipulation: intuitively, Billy is not responsible for his actions under the chip's influence. This type of case is a problem for compatibilists,[2] but easy for libertarians to deal with: if Billy's mother's manipulation is effective, it must be in virtue of removing his ability to do otherwise than the chip directs — and this ability is exactly what is required for libertarian responsibility.

This essay will argue that libertarians should worry about manipulation, too. To the extent that libertarians are seen to have a problem with cases of manipulation, discussion of the problem usually centres on Frankfurt cases. These are cases where the protagonist is supposed to lack moral responsibility according to libertarian criteria, but is

intuitively responsible for his actions. Many have found Frankfurt cases convincing,[3] but Widerker (1995) provides a powerful response on behalf of the libertarian.[4] The main part of this essay (§§1-4) will be taken up with a new sort of example, making the opposite type of problem for libertarians: this will be an example of an agent who *does* meet libertarian criteria for responsibility, but is *not* intuitively responsible. §1 presents the example; §2 answers the objection that the example cannot work as described; §3 looks at an analogy with the debate between hard and soft compatibilists; and §4 argues that compatibilist theory (more specifically, Fischer-Ravizza-style compatibilism) can yield the correct verdict on the example.

This case on its own does not constitute an insuperable obstacle for the libertarian; but the most promising libertarian response to the Smith/Brown case (this paper's centerpiece, described in §1 below) contradicts Widerker's response to Frankfurt cases. This poses a dilemma, as is argued in §5: either libertarian criteria for moral responsibility are too strong, in which case Frankfurt agents are not judged responsible, or they are too weak, in which case Smith is judged responsible.

1. Brown is a libertarian mad scientist. Brown is convinced that his fellow citizens are morally corrupt and need to be punished for it, and so he builds a ray-gun to wipe us all out. But Brown is a PR-conscious mad scientist, and he wants to be able to show the rest of the world that we deserved our fate. In particular, Brown wants to show that, given the opportunity, we would freely buy eggs from battery-caged hens rather than spend a few dollars more for certified organic eggs.[5] And so Brown abducts Smith (a typical person) in her sleep, transferring her to a simulation in his laboratory. When Brown activates the simulation, Smith finds herself (as far as she can tell) standing in front of the

egg display at the grocery store, with an empty cart, holding a shopping list that reads "Eggs." Now Brown wants to record evidence that Smith freely chooses to buy bad eggs; but as a libertarian, he knows that the very freedom of Smith's choice entails that it is possible she will choose the better ones instead. It seems, therefore, that Brown cannot be sure he will get the result he wants, and must abduct others until he gets the desired outcome.

But Brown is impatient to begin wiping people out, and cleverly devises a way to ensure he will get the desired outcome from Smith without interfering with her libertarian free will. If Smith chooses the better eggs, Brown uses his patented Memory Eraser[6] on her and resets his simulation, so that she will be faced with *exactly the same choice again*: once more, Smith finds herself standing in front of the eggs with an empty cart and a shopping list that reads "Eggs." In virtue of her libertarian freedom, there is no guarantee that she will make the same choice she did the first time around, just as Brown had no guarantee beforehand that she would make the choice he desired. Now, if Smith persists in choosing the better eggs, Brown will continue wiping her memory and resetting the simulation. Supposing Brown is clever enough to have invented an immortality drug, or some other means of ensuring that this sequence of choices can be continued indefinitely,[7] Smith is bound to choose the bad eggs eventually. For if it is genuinely possible for her to choose them over the better eggs, there must be a non-zero probability[8] of this outcome being realized at each instance of her being faced with the choice; and if a probabilistic process is repeated indefinitely, any outcome with a non-zero probability will eventually obtain. Thus, Brown is able to ensure that he will eventually get the outcome he desires, without interfering with Smith's ability to choose

otherwise: if Smith chooses the bad eggs at t_1, then it was possible for her to choose the better eggs at t_1.

At first blush, it may seem plausible to simply consider Smith responsible for her action at every iteration of the simulation, including the final one. After all, she does choose all by herself what to do at each stage. But this response becomes implausible after one final modification to the case as described. Suppose Brown has his simulation rigged up so that when Smith chooses to buy conventional eggs, she actually does buy them (in the real world, not just in the simulation), but when she chooses to buy organic eggs, she does not actually buy them. One way of spelling this out has it that the simulation is linked to an online grocery service. If Brown does not specifically interfere, then whatever Smith chooses to buy in the simulation she will actually buy from the grocery service; but whenever Smith chooses the better eggs, Brown *does* interfere, preventing Smith's choice from leading to action. With this set-up, if we take Smith to be responsible for all of her choices, we must take her to be responsible for actually buying conventional eggs, but not for buying organic eggs, despite the fact that she had no alternative but to actually buy the conventional ones. We can hardly consider her responsible for merely potential consequences of her earlier choices.

To be clear: it is a desideratum of theories of moral responsibility that they should not count Smith responsible for her final purchase, which will inevitably be of bad eggs. As §5 will argue, it is much more difficult for plausible libertarian theories to meet this condition than it is for plausible compatibilist theories.

2. But first, let us consider the response that the example cannot work as described: Brown's technique does not guarantee that Smith will eventually choose to buy the bad

eggs. The crucial premise in the argument that Brown's technique will guarantee the results he desires is this:

(PROB) If *S* is free (in the libertarian sense)[9] to do *A* in circumstances *C*[10],

then there is a non-zero objective probability[11] that *S* will do *A* in *C*.

The probabilities concerned here are hypothetical limiting frequencies. To say the probability of *S* doing *A* in *C* is *x* is to say that if *S* is repeatedly placed in circumstances *C*, then the ratio of instances where *S* does *A* to the total number of repetitions will approach *x* as the number of repetitions increases. It follows that if there is a non-zero probability of *S* doing *A* in *C*, then if *S* is repeatedly placed in *C* sufficiently often, *S* will eventually do *A*.

This paper does not offer a direct argument for (PROB). But to argue against the possibility of cases like Smith/Brown by denying (PROB) requires more than simply arguing that (PROB) is false, for (PROB) is stronger than necessary for setting up the Smith/Brown case: where (PROB) refers to an arbitrary action *A*, the Smith/Brown case only requires (PROB-egg), which substitutes purchasing bad eggs for *A* in (PROB). Now (PROB) may be false, while some instantiation (PROB-*B*) is true, for some specific action *B*; but in that case, we can revise the Smith/Brown case to have Brown trying to get Smith to freely *B* instead of freely buying bad eggs. The revised case would present the same problem of manipulation.

So the only reasonable way to object to the Smith/Brown case by attacking (PROB) would be to argue for a strong principle like the following:

(~PROB) If S freely (in the libertarian sense) chooses to do *A* in circumstances *C*, then *whenever S* is in *C*, *S* will choose to do *A*.

In order for this principle to make sense, "whenever" must be understood as an ordinary temporal "whenever"—quantifying over instantiations of *C* within a given world, not across worlds. The hypothesis of (~PROB) has it that *S* is free in the libertarian sense to do *A* in circumstances *C*, which must mean that there is a possible world where *S* does *A* in *C*, and a possible world where *S* refrains from doing *A* in *C*. So it seems that (~PROB) asserts either a remarkable uniqueness of the actual world, or quite a strong modal claim: if it is possible for *S* to avoid doing *A* in *C*, but *S* does *A* in *C*, then it is thenceforth impossible for *S* to avoid doing *A* in *C*, should *S* find herself in *C* again. That is, we can either interpret (~PROB) as asserting that *in our world*, it so happens that if one freely performs an action in given circumstances, then one will always do the same whenever those circumstances obtain again; or we can interpret it as asserting that *there are no worlds* where an agent performs an action in given circumstances once, but avoids performing it in those same circumstances at a later time. Both of these are strong metaphysical claims, and it is not obvious how one might justify them.[12,13]

(PROB), on the other hand, seems plausible, at least *prima facie*. It resembles an independence condition on repeated trials of circumstances *C*.[14] According to (PROB), what happens in the first trial ought not to determine the outcome of successive trials. If those circumstances *C* can be specified so that there is no significant causal link open between successive trials of *C*, it is indeed hard to see how the outcome of the first trial could affect, much less determine, the outcomes of later trials. And the Smith/Brown case we have been considering does provide circumstances with just this property: Smith

remembers nothing of previous iterations of the simulation; she is presented, in the simulation, with just the same situation at each iteration; her reasoning faculties are otherwise untouched; and Brown exerts no further influence on her choice. These are repeatable circumstances, and there is no clear way for Smith's first choice to influence her future choices.

3. So let us agree that Brown's technique is effective, and ask what this means for the compatibilist and the libertarian. This seems to be a genuine case of responsibility-undermining manipulation, that our best theory of responsibility should not consider Smith responsible for her final action. For although Smith's choice is free at each iteration of Brown's simulation—he does not directly interfere with her decision-making—she ultimately has no alternative but to buy the bad eggs. To put the intuitive point another, more compatibilist-friendly, way: Smith arguably does not have control over her final purchase, and so cannot be responsible for it. This is because there is only one action that can issue from her choices—regardless of what she chooses, she will only buy conventional eggs. Therefore, in a sense, the causal chain resulting in her purchase goes around her. Thus, it seems (at least *prima facie*) that defending a theory that counts Smith as responsible for buying bad eggs is to bite a bullet. Now, although the Smith/Brown case has been presented as a problem for libertarians, talk of biting the bullet in a case of manipulation inevitably raises thoughts of hard and soft compatibilism; so let us say just a few general words about that debate before considering how libertarians and compatibilists should deal with the Smith/Brown example specifically.

The terms "hard compatibilism" and "soft compatibilism" come from Kane (1996: 67-8). How hard a compatibilist one is depends on how one responds to cases of

covert manipulation like the one described in the introduction to this paper: soft compatibilists try to find a compatibilist theory that explains why the victims of such covert manipulation are not morally responsible for their resulting actions; while hard compatibilists "bite the bullet" and insist that contrary to our intuitions, those victims are indeed responsible. Fischer and Ravizza (1998), then, gives a soft compatibilist account of responsibility; while Watson (1999) and Russell (forthcoming) advocate hard compatibilism for general reasons.

This paper will not give a detailed discussion of the arguments on either side of this debate—that would take us too far afield. What we shall consider here is whether an analogue of hard compatibilism would make an adequate response for the libertarian to the Smith/Brown example. The first thing to note, of course, is that there is a reason we use the phrase "biting the bullet" rather than, say, "chewing the chocolate" for accepting the counterintuitive verdict on the example: there is a cost to contradicting our intuitions; it should be avoided if possible.[15] This is why, *ceteris paribus*, a soft compatibilist theory would be preferable to a hard compatibilist theory. Watson and Russell do not dispute this; what they argue is that we cannot come up with an adequate soft compatibilist theory without *ad hoc* conditions to rule out covert manipulation. If they are correct that soft compatibilism is not a viable option, then biting the bullet may be a viable strategy for the libertarian, if it can be shown that hard compatibilists must bite harder on cases of covert manipulation than libertarians would on the Smith/Brown example. But if, as Fischer (2004) argues, soft compatibilism is a live option, then the situation is different. Then, if compatibilists can handle the Smith/Brown case more easily than libertarians can, then the libertarian cannot simply bite the Smith/Brown bullet. Unless libertarians

already have a clear advantage over compatibilists prior to this paper, biting the bullet on Smith/Brown would mean conceding that (soft) compatibilists manage to get our intuitions on manipulation cases right, but libertarians do not. Let us, then, move on to look at how well compatibilists and libertarians can deal with the Smith/Brown example.

4. The challenge, recall, is to explain why Smith is not responsible for her purchase of bad eggs. It will be easier (though not easy!) for the compatibilist[16] to meet this challenge than for the libertarian, if only because the more sophisticated compatibilist accounts of moral responsibility tend to be shaped by trying to deal with difficult cases of manipulation like this one. Libertarians, on the other hand, have not had much need for innovation, since requiring alternative possibilities is such a powerful tool for fending off mad scientists. Here, that simple tool just doesn't work. Smith is never faced with a choice where she has but one possible option.

Compatibilists have more tools in their box. For example, one of the new ideas in the Fischer-Ravizza reasons-responsiveness account mentioned earlier is *mechanism ownership* (Fischer and Ravizza 1998: ch. 8, esp. §VIII). On this account, an agent acts freely only if the mechanism producing his action is his; he must have *taken responsibility* for the mechanism. Among the requirements for taking responsibility for a mechanism are seeing oneself as the source of the behaviour thus produced, and basing this belief on evidence. But, the compatibilist might argue, Smith lacks crucial evidence about the mechanism producing her egg purchase. She may well take responsibility for the ordinary sort of practical reasoning at play in each iteration of Brown's simulation, but she does not know that that mechanism is but a part of what our compatibilist claims is the mechanism ultimately producing the purchase of bad eggs. The mechanism

producing that behaviour might be described as *ordinary practical reasoning repeated until it yields the desired result*. If this is the mechanism at work, then Smith lacks any evidence that it is (thanks to Brown's Memory Eraser), making it impossible for her to genuinely take responsibility for her actions. The mechanism producing her behaviour is not hers.[17] Note that this response matches nicely with the intuitive reasons for taking Smith not to be responsible: it is up to Brown, not Smith, what sort of eggs she buys.

Todd Long, in comments on a presentation of this paper at the 2008 Pacific APA, claims that "Smith's practical reasoning repeated until it yields the desired result" cannot be a correct description of the mechanism that produces Smith's purchase. On the contrary, he argues, the mechanism producing Smith's purchase is just Smith's ordinary practical reasoning. The difference between the Smith/Brown example and an ordinary case of practical reasoning is in the inputs to this mechanism; which, in the case under discussion, includes Brown's influence. In other words, we might identify Brown's influence on Smith as determining the inputs to her practical reasoning—in particular, he causes her, over and over again, to reason as though she is in a grocery store needing to buy eggs, and so forth. Brown does not interfere further with her practical reasoning; Smith lacks evidence about where the inputs to her action-producing mechanism come from, not about what that mechanism is. She presumably takes ownership of her ordinary practical reasoning, so she must be responsible for buying the bad eggs. Therefore, according to Long, Fischer and Ravizza must take Smith to be responsible for her purchase.[18] Clearly, then, it is important for Fischer-Ravizza-style compatibilists to argue that the mechanism that yields Smith's purchase of bad eggs is not simply her ordinary practical reasoning.

Since Fischer and Ravizza do not give instructions for determining what mechanism is operating in any particular case, we can give no more than preliminary remarks here; but here is the beginning of a response to Long. First, note that if Long is right, then Smith's purchase is produced by the same mechanism as would be operating without Brown's influence. But if we take the outputs of the mechanisms relevant to responsibility to be *actions* rather than *choices* (as Fischer and Ravizza do—see, e.g., Fischer and Ravizza 1998: 74), then this identity is troubling. Smith's practical reasoning, absent Brown's influence, takes the same kinds of inputs as the mechanism in the Smith/Brown case as described; but only the former can yield the purchase of the better eggs as an output. To be sure, either mechanism can produce the *choice* to buy the better eggs, but in the Smith/Brown case, this choice cannot result in action. Our compatibilist can claim that although it is Smith's ordinary practical reasoning that yields her choice at each iteration of the simulation, the mechanism yielding Smith's eventual purchase differs from her ordinary practical reasoning because, thanks to Brown's influence, one of the possible outputs of her practical reasoning (viz., purchasing better eggs) is blocked.

To put the point another way, Brown's influence on Smith has two parts: he causes her practical reasoning to be fed the same inputs over and over again, and he causes her practical reasoning to result in action only if it results in the choice he wants. It is this second part of Brown's influence that Long's objection misses. This gives reason to think that although Brown has not interfered with Smith's practical reasoning, he has nevertheless interfered with the mechanism resulting in her action. It may be helpful in this connection to think of mechanisms as functions or rules assigning certain output-actions to certain input-reasons. If this is the right way to think about mechanisms, then

Brown clearly changes which mechanism is active, for he changes the facts about which actions result from which reasons—though, of course, he does not change the facts about which choices result from which reasons.

5. There is no promising libertarian strategy analogous to denying that Smith owns the mechanism producing her action. If we suppose that what matters for responsibility is the possibility of *choosing* otherwise (which, we shall see, is precisely what matters on Widerker's view), we may well require that the source of that possibility be in some sense located within the agent herself (so, for example, the alternative choice should not be possible because of some inherent randomness in one's decision-making processes), but it does not make sense to suppose that Brown's interference relocates the source of the relevant possibility outside of Smith herself. The metaphysics of Smith's choice at each iteration of the simulation is exactly the same as it would be without Brown's influence.

Note that what we might call a "minimally libertarian" version of the mechanism-ownership response might work, but would be difficult to motivate. That is, we could add to the mechanism-ownership response a bare requirement that responsible agents not be causally determined to act as they do. However, this would leave open the question of why we should bother with the libertarian condition, having already accepted the full machinery of Fischer and Ravizza's compatibilism. Or, to worry in a similar way but from the opposite direction: having accepted that responsible agents must be undetermined, do we really need to appeal to such a complex theory as Fischer and Ravizza's just to deal with a relatively small group of examples? At the very least, the compatibilist version of the mechanism-ownership account has the advantage of

simplicity over its minimally libertarian counterpart.

In fact, the best libertarian strategy for dealing with the Smith/Brown case, like the compatibilist strategy highlighting mechanism ownership, also follows the intuitive reasons for taking Smith not to be responsible. The intuitively salient point about Smith's alternative possibilities in the case described is not that she has many options at each iteration, but rather that she has only one possible final outcome. Despite her ability to choose-at-t_1 otherwise than she did, Smith is not able to choose-at-the-end-of-the-sequence otherwise than she did. It is Brown, not Smith, who determines what happens at the end of the sequence; if there are alternative possible endings to the sequence, it is due to Brown, and so it is Brown who should be judged morally responsible for Smith's purchase at the end of the sequence.

But if this distinction among modalities is to be more than an *ad hoc* solution to the specific problem at hand, we will need a broader principle for determining what sort of ability-to-do-otherwise an agent needs in order to act freely. Now we come to the dilemma hinted at in the introduction. For it seems that any general principle telling us to look at the sense in which Smith could not do otherwise must tell us to worry about the possibility of an agent's choice having a different *outcome* than it in fact does, rather than simply about the possibility her *choosing* otherwise than she does. That is, it is difficult to see how we can rule that Smith is not responsible on the basis of the impossibility of her doing otherwise unless we claim that what matters is the fact that even if she had chosen to buy better eggs at t_1, she still would eventually wind up choosing (at some later t_2) to buy bad eggs. But this makes Widerker's objection to Frankfurt cases unavailable. To see this, we must describe a Frankfurt case, and briefly explain Widerker's objection.

Black is a Republican mad scientist, and there is an election in progress.[19] Black believes he owes his party more than just one vote, and so yesterday he implanted Jones with a microchip to ensure that Jones also votes Republican today. But though Black is mad, he is also careful, and prefers not to interfere with Jones's mind more than necessary, for fear of detection. Thus, the implanted chip is a mere fail-safe device: if Jones would have voted Republican without Black's interference, the chip will do nothing; but if Jones would otherwise vote Democrat, the chip is activated, and compels Jones to vote Republican instead. As it happens, Jones decides on his own to vote Republican, and the chip is never activated. It seems, then, that Jones should be judged responsible for his vote, since he acted as he did without Black's influence. (We may assume that the process of implanting Jones with the chip did not itself affect Jones's decision.) But Jones could not have done otherwise. Since it was impossible for him not to vote Republican, it seems the libertarian must not consider him responsible for his vote.

Widerker's reply—which appears to be the best available to the libertarian[20]—has us re-examine the action with respect to which we are to judge Jones's responsibility. For the libertarian, he insists, "mental acts such as deciding, choosing, undertaking, forming an intention… constitute the basic *loci* of moral responsibility" (1995: 247). So in evaluating Jones's responsibility, we must ask not whether it was possible for him not to vote Republican, but whether it was possible for him not to *choose* to vote Republican. It is difficult to see how Black could devise a microchip to act in the desired way with regard to such a metaphysically singular action, as opposed to a complex one such as voting. The chip must be designed either to be activated before Jones's decision, or to be

activated after the decision. If it is designed to act before Jones's decision, there must be some cue which will tell the chip whether to activate or not. That is, there must be some sign S such that if S occurs at t_1, then at some later t_2, Jones will decide not to vote Republican. But if there is such a sign, then Jones's decision at t_2 is causally determined by a prior state of the world. To insist that this decision is free is simply to beg the question against the libertarian. On the other hand, if the chip is to be activated after the decision, then it does not genuinely remove the possibility of deciding not to vote Republican. If Jones decides not to vote Republican, he will be overridden by the chip before he can act on that decision—but he has made the decision. The libertarian can certainly account for Jones's freedom in this case.

But this response runs directly contrary to what was above called the most promising libertarian account of the Smith/Brown case. To account for Smith's lack of responsibility, it was recommended that the libertarian focus on the fact that she has no alternative but to *act* as she did, despite the face that she could have *chosen* otherwise. Widerker's response to Frankfurt cases says that Jones is responsible because he has the possibility of *deciding* to vote Democrat—if only for a short time, before Black overrides that decision—despite the fact that Black removes his ability to *act* otherwise. If Widerker-style libertarians are to hold that Jones is responsible for his vote, then, it seems they must also hold that Smith is responsible for her eventual buying of eggs in Brown's simulation. The libertarian must bite the bullet and accept the counterintuitive upshot of one of the two types of cases: either Smith is responsible or Jones is not.[21]

[1] Or at least, this is the case for compatibilists in the mold of Fischer and Ravizza (1998), which will be the brand of compatibilism primarily under discussion in this paper. Compatibilists of other kinds may attend to other aspects of Billy's situation instead of the mechanism producing his actions, but it is generally agreed that cases of this kind make trouble for compatibilists. Thanks to an anonymous referee here.

[2] This is not necessarily an *insoluble* problem — see Fischer (2004) for an effort to respond to the criticism as directed at his and Ravizza's account in particular, and Double (1989) for a defence of reasons-responsiveness theories in general.
[3] E.g., Fischer and Ravizza (1998); Dennett (1984); Zimmerman (1988).
[4] Versions of Widerker's reply are also endorsed, notably, by Ginet (1997) and Kane (1996). Similar objections to Frankfurt cases can be found in Kane (1985). Though there are already several responses to Widerker's reply in the literature (see note 20), these three prominent libertarians remain unmoved; that is why this paper attempts a new angle of attack on Widerker's argument.
[5] We may assume that where Smith lives, organic standards are well-enforced, and include robust animal welfare provisions. In short, let us assume that buying conventional eggs is, for Smith, morally worse than buying organic eggs. Henceforth, for clarity, let us simply refer to the conventional and organic eggs as "bad" and "better", respectively.
[6] Brown developed this device while working with another abductee, Sleeping Beauty, during experiments on subjective probability (Elga 2000).
[7] Readers whose science-fictional imagination is stretched too far by such a supposition may wish to replace Brown with God, which leads to some other interesting considerations. See also notes 12 and 15.
[8] That Smith's decision is describable probabilistically should not be taken to imply that it is *random* in any sense which would interfere with its freedom. When a basketball player of a certain skill level shoots a free throw, there is a certain probability that the shot will go in, but whether the shot goes in is (in ideal circumstances—say, the player is alone in a gym with negligible air resistance) entirely determined by the player's ability to correctly control his muscles. The event can be described probabilistically, but the responsibility for a make or a miss belongs only to the player herself.

Contrastingly, Peter van Inwagen (2000: 13-18) describes a case similar to mine—an agent is presented with exactly the same choice, in exactly the same circumstances, over and over again—to make the point that if indeterminism is true in the way required for the consequence argument against compatibilism, then our apparent choices appear to be mere chance. But despite the similarity of our cases, consideration of van Inwagen's conclusions would be outside the scope of this paper.
[9] Some libertarians, such as Kane (1996), hold that there are actions which can be performed freely despite being determined. What matters for these libertarians is that one's character be formed by actions which could have been otherwise (cf. Kane's "self-forming actions"). A predetermined act may yet be free if one's freely-formed character is what determines the act. To run our Smith/Brown example for a Kane-type libertarian, suppose that instead of putting her in a situation where she must buy organic or conventional eggs, Smith must perform some self-forming action. The rest of the argument runs the same as in the text.
[10] The circumstances referred to here should not be understood as a complete description of the universe at the time of S's choice, for these circumstances are to be repeatable. Rather, what we should have here is a complete description of the causal factors relevant to S's choice. In the Smith/Brown case, anything outside Smith and the simulation will be irrelevant to Smith's choice of eggs.
[11] There are cases where it seems plausible to say that some outcome is possible, but has probability zero. For example, if one takes a real number at random from the interval (0,1), one has probability zero of choosing a rational number—but this is certainly a possible outcome. However, it is not clear that such cases need touch the issue at hand, for they generally rely on an infinite space of possible outcomes. It is enough for present purposes that Brown be able to get Smith to buy the appropriate eggs; he need not be able to get her to do so in an infinitely precise way.
[12] Molinists and anti-Molinists may have a particular interest in (PROB) and (~PROB), since these two principles have obvious connections to counterfactuals of freedom. An endorsement of one principle or the other would constrain one's answer to the question of what grounds the truth of a counterfactual of freedom like "If *S* is in *C*, *S* will do *A*." And, *vice versa*, if one has a settled opinion of what makes counterfactuals of freedom true, one might be able to derive therefrom (PROB) or (~PROB), or something in between the two.
[13] One further objection to the effectiveness of Brown's technique: on Robert Kane's account, the intentions involved in a free action must be revisable. That is, if Smith freely forms an intention to buy bad eggs, it must be possible for her to change her mind right up to the point when she actually buys the eggs. If, to guarantee that Smith buys bad eggs, Brown must fix her intention to do so, once it is formed, making it irrevisable, then Kane's theory can explain why Smith is not responsible for buying bad eggs.

But Brown need not fix Smith's intention in order to guarantee his desired outcome. For if Smith does revise her intention to buy bad eggs before consummating the act, Brown can use his Memory Eraser and reset the simulation, just as if Smith had not formed the intention to buy bad eggs in the first place. Unless revising the intention to buy bad eggs means *immediately* making good on an intention to buy better eggs instead, Brown has a chance to reset the simulation before Smith does what he does not want her to do. If he is patient, Brown can keep resetting the simulation until Smith finally forms *and keeps* an intention to buy bad eggs.

[14] Thanks to Paul Bartha for pointing this out.

[15] Free will theodicists, moreover, have a special incentive not to bite the bullet. For a central point in such theodicies is that God values libertarian free will so highly that it is better for Him to allow us to do evil freely than to override that freedom and compel us to do good. See, e.g., Plantinga (1974: 42-44). But if Smith, in the Smith/Brown example, does act freely, then it seems Brown has found a recipe God could use to guarantee that we freely do good: keep re-creating the world until we make the right choices every time. Of course, God has some moral scruples that Brown does not, so some argument would be needed before we could write off free will theodicies entirely; but if such an argument can be made, free will theodicists must refine their notion of libertarian free will so as to rule out examples like mine.

[16] Or at least, the soft compatibilist. Hard compatibilists may be less sympathetic to what I say on behalf of "the compatibilist" here. Thanks to an anonymous referee for making this clear to me.

[17] In the example as described, Smith's mechanism is also, arguably, not reasons-responsive. But this is a poor diagnosis of the reason we do not consider her responsible, for we could have a case where the manipulator's choice of the ultimate result (the halting conditions for the simulation, so to speak) *is* reasons-responsive in an appropriate way.

[18] But Long does not think this is a problem for Fischer and Ravizza; he thinks that it is entirely correct to take Smith to be responsible in this case, and indeed so to consider the agents in a wide range of manipulation cases. See Long (2004).

[19] The following is paraphrased from Frankfurt (1969: 835).

[20] This is not to say that it is universally accepted as decisively dealing with Frankfurt cases, of course. Widerker's response has itself generated several responses, e.g., Mele and Robb (1998, 2003), Fischer (1999, 2001), Hunt (2000), McKenna (2003). For responses to these responses, see, e.g., Ginet (2003), Goetz (2005), Kane (2000, 2003), Widerker (2000).

[21] Thanks to Murat Aydede, Paul Bartha, Tim Christie, Oisín Deery, Josh Johnston, Robert Kane, Todd Long, Alirio Rosales, Margaret Schabas, Max Weiss, audiences at the UBC Graduate Colloquium and the 2008 Pacific APA, and especially to Paul Russell and an anonymous referee for this journal for helpful comments on earlier versions of this paper.

References

Double, Richard (1989). "Puppeteers, Hypnotists, and Neurosurgeons," *Philosophical Studies* 56: 163-173.

Dennett, Daniel (1984). *Elbow Room: Varieties of Free Will Worth Wanting*. Cambridge: MIT Press.

Elga, Adam (2000). "Self-Locating Belief and the Sleeping Beauty Problem," *Analysis* 62: 292-296.

Fischer, John Martin (2004). "Responsibility and Manipulation," *Journal of Ethics* 8: 145-177.

Fischer, John Martin and Ravizza, Mark (1998). *Responsibility and Control: A Theory of Moral Responsibility*. Cambridge: Cambridge UP.

Frankfurt, Harry (1969). "Alternate Possibilities and Moral Responsibility," *Journal of Philosophy* 66: 829-839.

Ginet, Carl (1996). "In Defense of the Principle of Alternative Possibilities: Why I Don't Find Frankfurt's Arguments Convincing," *Philosophical Perspectives* 10: 403-417.

Ginet, Carl (2003). "Libertarianism," in Michael Loux and Dean Zimmerman (eds.), *Oxford Handbook of Metaphysics*. Oxford: Oxford UP, pp. 587-612.

Goetz, Stewart (2005). "Frankfurt-Style Counterexamples and Begging the Question," *Midwest Studies in Philosophy* 29: 83-105.

Hunt, David (2000). "Moral Responsibility and Unavoidable Action," *Philosophical Studies* 97: 195-227.

Kane, Robert (1985). *Free Will and Values*. Albany, NY: SUNY Press.

Kane, Robert (1996). *The Significance of Free Will*. Oxford: Oxford UP.

Kane, Robert (2000). "Responses to Bernard Berofsky, John Martin Fischer, and Galen Strawson," *Philosophy and Phenomenological Research* 50: 157-167.

Kane, Robert (2003). "Responsibility and Frankfurt-Style Cases: A Response to Mele and Robb," in Widerker and McKenna (2003), pp. 91-105.

Long, Todd R. (2004). "Moderate Reasons-Responsiveness, Moral Responsibility, and Manipulation," in Joseph Campbell, Michael O'Rourke, and David Shier (eds.), *Freedom and Determinism*. Cambridge: MIT Press, pp. 151-172.

McKenna, Michael (2003). "Robustness, Control, and the Demand for Morally Significant Alternatives: Frankfurt Examples with Oodles and Oodles of Alternatives," in Widerker and McKenna (2003), pp. 201-217.

Mele, Alfred R. and Robb, David (1998). "Rescuing Frankfurt-Style Cases," *Philosophical Review* 107: 97-112.

Mele, Alfred R. and Robb, David (2003). "Bbs, Magnets and Seesaws: The Metaphysics of Frankfurt-Style Cases," in Widerker and McKenna (2003), pp. 127-138.

Plantinga, Alvin (1974). *God, Freedom, and Evil*. Grand Rapids: Wm. B. Eerdmans Publishing Co.

Russell, Paul (forthcoming). "Selective Hard Compatibilism," in Joseph Campbell, Michael O'Rourke, and Harry Silverstein (eds.), *Action, Ethics and Responsibility: Topics in Contemporary Philosophy*, Vol. 7. Cambridge: MIT Press.

van Inwagen, Peter (2000). "Free Will Remains a Mystery," *Philosophical Perspectives*

14: 1-19.
Watson, Gary (1999). “Soft Libertarianism and Hard Compatibilism,” *Journal of Ethics* 3: 351-365.
Widerker, David (1995). “Libertarianism and Frankfurt’s Attack on the Principle of Alternative Possibilities,” *Philosophical Review* 104: 247-261.
Widerker, David (2000). “Frankfurt’s Attack on the Principle of Alternative Possibilities: A Further Look,” *Philosophical Perspectives* 14: 181-201.
Widerker, David and McKenna, Michael (eds.) (2003). *Moral Responsibility and Alternative Possibilities*. Aldershot, UK: Ashgate Press.
Zimmerman, Michael J. (1988). *An Essay on Moral Responsibility*. Totowa, NJ: Rowman and Littlefield.