

*Selective Hard Compatibilism**

Paul Russell

University of British Columbia

August 2007

Forthcoming in Joseph Campbell, Michael O'Rourke and Harry Silverstein, eds.,
Action, Ethics and Responsibility: Topics in Contemporary Philosophy, Vol. 7
(Cambridge, Mass.: MIT Press, forthcoming).

I. Compatibilism , Implantation and Covert Control

Recent work in compatibilist theory has focused a considerable amount of attention on the question of the nature of the capacities required for freedom and moral responsibility. Compatibilists, obviously, reject the suggestion that these capacities involve an ability to act otherwise in the same circumstances. That is, these capacities do not provide for any sort of libertarian, categorical free will. The difficulty, therefore, is to describe some plausible alternative theory that is richer and more satisfying than the classical compatibilist view that freedom is simply a matter of being able to do as one pleases or act according to the determination of one's own will. Many of the most influential contemporary compatibilist theorists have placed emphasis on developing some account of "rational self-control" or "reason-responsiveness".¹ The basic idea in theories of this kind is that free and responsible agents are capable of acting according to available reasons. Responsibility agency, therefore, is a function of a general ability to be guided by reasons or practical rationality. This is a view that has considerable attraction since it is able to account for intuitive and fundamental distinctions between humans and animals, adults and children, the sane and the insane, in respect of the issue of freedom and responsibility. This an area where the classical account plainly fails.

In general terms, rational self-control or reasons-responsive views have two key components. The first is that a rational agent must be able to recognize the reasons that are available or present to her situation. The second is that an agent must be able to "translate" those (recognized) reasons into decisions and choices that guide her conduct. In other words, the agent must not only be aware of what reasons there are, she must also be capable of being moved by them. This leaves, of course, a number of significant problems to be solved. For example, any adequate theory of this kind needs to be able to explain just how strict and demanding this standard of practical rationality is supposed to be. On the one hand, it is clearly too demanding to insist that agents must always be able to be guided by available reasons – otherwise an agent could never be held responsible for failing to be guided by the available reasons. On the other hand,

more is required than that the agent is occasionally or intermittently guided by her reasons. An agent of this kind is not reliably and regularly rational to qualify as a free and responsible agent. So some set of conditions needs to be found that avoids both these extremes. This is not, however, the problem that I am now concerned with.²

Let us assume, with the proponents of compatibilist theories of rational self-control, that there is some account of these capacities that satisfy these various demands. We may call these the agent's RA capacities, as they provide for rational agency. This account still faces another important set of problems as presented by incompatibilist critics. One famous problem with classical compatibilist accounts of moral freedom ("doing as we please") is that agents of this kind could be manipulated and covertly controlled by other agents and yet, given the classical compatibilist account, still be judged free and responsible. This is, as the critics argue, plainly counter-intuitive. Agents of this kind would be mere "puppets", "robots" or "zombies" who are "compelled" to obey the will of their covert controllers. Agents of this kind have no will of their own. They are not real or genuine agents. They only have the façade of being autonomous agents. When we discover the origins of their desires and willings – located with some other controlling agent – then our view of these (apparent) agents must change. The deeper problem with these (pseudo) agents, incompatibilists argue, is that while they may be "doing as they please" they have no control over their own will (i.e. they cannot shape or determine their own will). It is, therefore, an especially important question whether compatibilist accounts of rational self-control can deal effectively with objections of this kind.³

Rational self-control theories have two ways of approaching this problem. The first is to argue that what troubles us in situations of this kind, where manipulation and covert control is taking place, is that the agent's capacity for rational self-control is in some way being impaired or interfered with. The process of brainwashing, neurological engineering, or some other form of mind-control operates by way of damaging the agent's capacity to recognize and/or respond to the relevant reasons that are available.⁴ The situation may, however,

be more subtle and complicated than this. The manipulation or covert control, incompatibilists argue, can also work without impairing the agent's rational capacities but by controlling the way that they are actually exercised in particular circumstances. This sort of case is much more problematic for the compatibilist. Per hypothesis, the agent continues to operate with the relevant rational dispositions (recognition, reactivity, etc.). The way that this capacity is actually exercised in particular circumstances – i.e. whether the agent's conduct succeeds or fails to track the available reasons – is not under the agent's control, as this would require libertarian free will to act otherwise in identical circumstances. In normal circumstances, the explanation for success or failure will rest with natural causes that involve no external controller or manipulation. The incompatibilist objection, however, is that there could be a situation whereby the agent is controlled in this way and the compatibilist has no principled reason for denying that the agent is free and responsible. It follows, therefore, that compatibilist accounts of rational self-control cannot provide an adequate account of conditions of free and responsible agency. Agents who are manipulated and covertly controlled are obviously not free and responsible despite the fact that they may possess a general capacity for rational self-control of the kind that compatibilists have described.

II. Libertarianism and the Implantation Standard

According to incompatibilists, the problem with manipulation and covert control is one that indicates a more fundamental and general weakness in the compatibilist position. Following some prominent compatibilist accounts, let us assume that our power of rational self-control presupposes that the agent possesses some relevant “mechanism” M, whereby the agent who possesses M is able to recognize and react to reasons.⁵ Incompatibilists argue that cases of implantation highlight aspects of compatibilism in relation to the way that M is acquired and operates that is problematic even when manipulation and covert control is absent (i.e. in the “normal case”). Let us assume that M may be “implanted” by natural, normal causal processes that involve no manipulation or

covert control by other agents. Implantation of M in these circumstances is “blind” and without any artificial interference of any kind. What still troubles us about these cases is that while the agent may be a rational self-controller, she nevertheless lacks any control over the way that these capacities are actually exercised in specific circumstances. Whether M is such that in conditions C the agent will succeed or fail to track the available reasons is not something that depends on the agent. Even in these “normal” cases the agent is still subject to luck regarding the way that M is actually exercised in C. In order to avoid this problem the agent must be able to choose or decide differently in the very same circumstances. An ability of this kind – let us call it exercise control - would require the falsity of determinism and some kind of libertarian free will. What ought to bother us about manipulation and covert control, therefore, is not simply that some other agent decides how the agent’s will is exercised in conditions C, but that the agent himself lacks any such ability or power. It is a matter of luck how his powers of rational self-control are actually exercised. Clearly, then, the lack of exercise control is not a problem that arises only in (abnormal or deviant) cases of manipulation and covert control.

It may be argued that one way that compatibilists will be able to avoid this difficulty, without collapsing into libertarian metaphysics, is to give more thought to the problem of how M is acquired by the agent. That is to say, since the agent is held responsible for the upshots that issue from M (in particular circumstances), it surely must follow that the agent has some control over how M is acquired. Failing this, the agent will lack control not only over the particular way M operates in C, but also over the fact that it is this particular mechanism M that he is operating with. Another mechanism, M#, may produce a different upshot in C. The agent, on the suggested compatibilist account, has control over none of this. What is need, therefore, in the absence of exercise control, is some control over mechanism acquisition. The incompatibilist will argue, however, that the compatibilist cannot provide any plausible account of how this could be possible.

While we can make good sense of having control over our actions on the basis of possessing some reasons-responsive mechanism (M), it is not at all obvious what it means to say that an agent controls the process of acquiring such mechanisms. The mechanisms that we acquire generally develop through a process of (moral) education that begins at a very early stage of life. For this reason responsibility for the kinds of mechanism that children acquire and develop rests more plausibly on the shoulders of the adults who have raised the child. Moreover, even at a later stage (e.g. adolescence) when a person becomes able to think critically about the way their own deliberative capacities actually operate, there is little or no question of the agent being able to radically modify or reform the mechanisms that he is (already) operating with. Control of this kind is not available even to mature adults, much less younger children.

Let us concede, nevertheless, that we can make some sense of the suggestion that the mature agent has control over mechanism acquisition. This form of control must itself depend on the agent's ability to deliberate and decide about mechanism selection on the basis of some mechanism that he already has. This situation presents compatibilist theory with a serious problem. The selection of some mechanism must be based on some mechanism that the agent currently operates with. This mechanism must be either chosen or given (i.e. through processes that the agent does not control). At some point, the mechanism involved in the process of mechanism acquisition must itself have been unchosen or presented to the agent through a process that he did not control (natural or artificial). Any choice concerning mechanism acquisition, therefore, must eventually depend on unchosen mechanisms – even on the optimistic assumption that mature agents are able to make choices of this kind.

It is evident that incompatibilist criticism of compatibilist theories of rational self-control reach well beyond narrow worries about manipulation and covert control. The deeper worries that situations of this kind bring to light is that rational self-control provides no (final) control over mechanism acquisition, nor over the way that these mechanisms are actually exercised in specific circumstances (i.e. success or failure to track reasons in particular conditions is

not open for the agent to decide). Given these criticisms, it follows that agents who operate with rational capacities of these kinds are subject to luck about what specific mechanism (M) they acquire and operate with, as well as luck about the way the mechanism they operate with is actually exercised. While these problems are certainly manifest in manipulation cases, they are by no means limited to them. On the contrary, these problems of limited control and luck are systematic to all circumstances in which agents operate on the compatibilist rational self-control model.⁶

Incompatibilist libertarians will be quick to contrast their own situation with respect to implantation and manipulation issues. Let us suppose that it is possible to artificially (intentionally) implant a mechanism of some kind that supports a libertarian capacity for free will. (Here again, the ontological basis of this mechanism is not our concern. There could be a biological basis for this or some soul-substance, and so on.) We may call a mechanism of this kind an ML-mechanism. The fact that ML is implanted by another agent (God, a neurosurgeon, etc.) will not trouble the incompatibilist, because implantation as such will not compromise the agent's ability to operate with control over the way that her rational capacities of deliberation and choice are actually exercised. Although the ML mechanism has been implanted by another agent, this does not make it possible to covertly control the implanted agent by means of this process. On the contrary, since the implanted agent possesses exercise control over the way this mechanism operates (i.e. controls the way reason move her), the source or historical origins of ML is irrelevant to the way that the agent chooses to exercise her ML powers in specific circumstances. So long as the agent is capable of exercise control – then it is possible to set aside the issue of mechanism acquisition as irrelevant because there is no threat of manipulation or covert control.

It is true, of course, that libertarian agents who operate with ML mechanisms, could not themselves control the process of ML acquisition unless they first possessed some ML mechanism. It follows from this that even libertarian agents of this kind do not control the process by which they become

capable of libertarian free will (this must be a “gift of nature” or “God-given, etc.). Nevertheless, as I have explained, whether the implantation process in this case involves natural (blind) processes or artificial (other agent) involvement is irrelevant. The nature of the mechanism implanted precludes manipulation and covert control. More importantly, it ensures that the agent is not simply “lucky” or “unlucky” in relation to the way reasons actually move her. Because the agent possesses an ML mechanism her will is truly her own. Her will is truly her own because (per hypothesis) she has the power to determine when and how reasons guide her conduct. That is to say, her will is truly her own not because she chooses to be a ML agent but because being an ML agent allows her to determine her own will.

The immediate significance of this libertarian response is that it serves to provide an alternative standard by which compatibilist theories may be judged. That is to say, the libertarian may argue that in the case of libertarian ML capacities it is possible to implant them without compromising the agent’s freedom and responsibility. This is a standard, the incompatibilist argues, that the compatibilist cannot meet. Compatibilists cannot meet this standard because the implantation of reason-responsive or rational self-control capacities (M) is consistent with the possibility of manipulation and covert control. Even if compatibilists reject worries about “luck” in relation to the way that these capacities are actually exercised, surely no compatibilist can allow that an agent is free and responsible in circumstances where she is being manipulated and covertly controlled by some other agent by means of some (abnormal) implantation process?⁷

III. Soft Compatibilism and History

Critics of libertarianism argue, as we know, that all efforts to make sense of libertarian powers that could deliver on exercise control run into problems of intelligibility and/or non-existence. Even if this is true, however, the compatibilist is still left with the problem of manipulation and covert control.

More specifically, compatibilists who rely on accounts of rational self-control need to take a stand on whether or not the presence of covert control and manipulation will necessarily compromise an agent's freedom and responsibility when it is clear that the agent's general powers of rational self-control are not impaired or damaged by this process. There are two different approaches that compatibilist may take to this issue. The first is that of the "soft compatibilist" who holds that the presence of manipulation or covert control by means of mechanism implantation rules out freedom and responsibility.⁸ The difficulty that soft compatibilists face is that, on the account provided, we could find ourselves with two agents (P / P*) who have identical moral capacities and properties and deliberate, decide and act in the very same way, and yet one is judged responsible and the other is not. The basis for this distinction rests entirely with the fact that one agent, P, has her deliberative mechanisms produced by natural (blind) causes, whereas the other agent, P*, has her deliberative mechanisms produced by some other agent. While we may concede that there is some residual intuitive worry about P*'s circumstances, the question is on what principled basis can the soft compatibilist draw such an important distinction? It will not suffice for soft compatibilists to simply assume that the contrast in causal origins matters and then to construct an ad hoc set of principles to rule out manipulation and covert control cases. This leaves their position vulnerable to the criticism that they have failed to identify the real root difficulty in their position (i.e. that these agents lack exercise control).

The most convincing soft compatibilist reply to this problem that I know of is provided by John Fischer and Mark Ravizza.⁹ What they argue is that reason-responsiveness or powers of rational self-control will not suffice for moral responsibility. This is because it is also necessary that agent's own the mechanism they are operating with. The problem of manipulation and covert control concerns the issue of ownership, not that of the agent's capacity for rational self-control. Briefly stated, an agent owns the deliberative mechanisms that issues in her conduct only if it has the right history or causal origins.¹⁰ When a mechanism is implanted by some other agent, using artificial techniques

of some kind, then “ownership” is compromised. On the other hand, when the mechanism is produced by means of normal causal processes, then ownership is not compromised. In the case of artificial implantation involving deviant causal processes, the problem is not that the agent’s rational self-control is compromised but that the agent does not *own* the mechanism that issues in her conduct.

The question we now face is does this appeal to ownership and history provide a secure basis for soft compatibilism? The first thing to be noted here is that there is no suggestion that ownership depends on control over mechanism acquisition or requires that the agent has somehow consented to the mechanism that she possesses and operates with. In fact, for reasons that we have already considered, any requirement of this kind is highly problematic and will inevitably run into regress difficulties. Clearly, then the distinction between acceptable and unacceptable processes of mechanism acquisition cannot depend on considerations of this kind. Nor is it obvious why one agent is said to own her own mechanism when it is naturally produced whereas the other does not because it has been artificially produced. In both cases the agents clearly possess these mechanisms and operate with them and in neither case have they consented or chosen their own mechanisms.¹¹

These considerations suggest that it remains unclear why any one should care about the different histories of mechanism acquisition when the “current time-slice” properties of both agents P/P* are exactly the same. From the perspective of both the agent herself, as well as those she is engaged with in her moral community, there is no difference at all between P/P*. There is no ability one has that the other does not also have, and both agents exercise these abilities in the exact same way. In other words, from both the internal and external perspective these two individuals are “inter-changeable” in respect of all powers and abilities that matter (per compatibilist hypothesis) to moral responsibility. In the absence of some further explanation for why “history” matters, therefore, the soft compatibilist way of dealing with manipulation and covert control cases seems arbitrary and ad hoc. Given that the soft compatibilist position depends

on placing weight on historical considerations we may conclude (for our present purposes) that the soft compatibilist strategy fails.

IV. Should Hard Compatibilists Just “Bite the Bullet”?

Where do these observations about contemporary compatibilist strategies that reply on accounts of rational self-control leave us? I think that it is clear that compatibilist accounts of this kind face the following dilemma. Either they must provide a more convincing account of why the history or causal origins of reason-responsive mechanisms matters OR they must accept that, since history is irrelevant, some version of hard compatibilism is the right course to take. I have explained that there is some reason to be sceptical about the prospects of the first alternative, so let us take a brief look at the hard compatibilist alternative. Robert Kane has noted that hard compatibilists are willing to “bite the bullet” and so deny that the presence or theoretical possibility of covert control or manipulation in any way compromises an agent’s freedom and responsibility.¹² This position certainly has some advantages when it comes to defending the compatibilist corner. One of these advantages is that it avoids a gap between libertarianism and compatibilism when it comes to the “implantation standard” that we considered above.

Recall that incompatibilist libertarians raised the following problem for compatibilist views. Compatibilist accounts of rational self-control exclude a power of exercise control and because of this it possible is for the relevant mechanisms to be implanted in the agent by some artificial means that would permit manipulation and covert control. Granted that manipulation and covert control compromise an agent’s freedom and responsibility, it follows that these compatibilist accounts fail to meet the implantation standard. In contrast with this, the implantation of libertarian deliberative mechanisms, which provide exercise control, will not leave any scope for manipulation or covert control by another agent. It is entirely irrelevant whether the libertarian mechanism is implanted by means of some natural, normal process or by some artificial

intervention by another agent. Implantation in this case does not make possible manipulation or covert control. This opens up a significant gap between the two views – one that soft compatibilists have tried to close (unsuccessfully) by appealing to history.

The hard compatibilist response is to deny the (incompatibilist and soft compatibilist) assumption that manipulation or covert control necessarily compromise freedom and responsibility. Their claim is that provided a suitably rich and robust account of moral capacity has been articulated and shown to be possible within compatibilist constraints (i.e. deterministic assumptions), then the mere fact that the agent may be covertly controlled or manipulated by this means is no more evidence that the agent is not free and responsible than it would be if natural, normal causal processes were at work and the source of the agent's deliberative mechanisms. In other words, the hard compatibilist runs the argument in reverse. If we can provide a suitable account of rational self-control, where the relevant mechanism is not implanted by some other agent or a deviant causal process, it follows (since origins are irrelevant to the functioning of this mechanism) that even if the mechanism has been implanted in a "deviant" manner that permits manipulation and covert control, there is no legitimate basis for denying that the agent is free and responsible. (E.g. If I discover this later evening that God, not nature, has implanted my deliberative mechanism and controls me through it, I still have no reason to change my fundamental conception of myself as a free and responsible agent. After all, I am unchanged and unaffected in all respects relating to my abilities, deliberations and conduct. There is nothing I was able to do then that I cannot do after being informed about the causal history of my deliberative mechanism.) If we opt for compatibilism, therefore, we must accept the hard compatibilist implications that go with it. If we can't live with this, then we need to turn to incompatibilism and/or libertarian metaphysics to avoid these worries about manipulation and covert control.

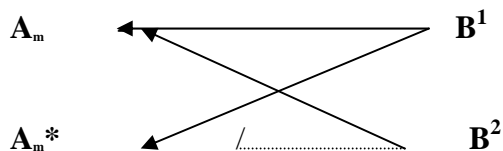
My view is that this is the right general strategy up to a point. However, I want to suggest a significant amendment or qualification to this hard

compatibilist alternative. Assuming that some suitably rich and robust account of moral capacity can be developed within compatibilist constraints, must we accept unqualified hard compatibilism? It is clear, I think, that there is something more to the basic intuition that covert control and manipulation compromise responsibility than the straight “bite the bullet” view allows for. It may be possible, however, to provide an alternative explanation for the source of our intuitive discomfort with this situation. We may begin by noting that the whole point of developing a theory of moral capacity is to describe the circumstances in which our moral sentiments of praise and blame, and the retributive practices associated with them, may be deemed appropriate or fair.¹³ The basic idea here is that the agent who is held responsible is a legitimate target of the moral sentiments of other members of her moral community in virtue of possessing the relevant set of capacities and abilities (i.e. rational self-control). Consider now some different scenarios that may arise in circumstances where deviant implantation and covert control is present.

Consider, first, an agent A_m with relevant (reason-responsive) moral capacity M . M is such that, bracketing off any worries about manipulation and covert control by others, it is reasonable and legitimate for another person B^1 to hold A_m responsible for the conduct that issues out of M . In other words, we assume some satisfactory compatibilist account of M , in circumstances where worries about manipulation by others do not arise. Now consider another scenario where agent A_m^* has the very same (time-slice) moral capacity M and the very same conduct issues from M . In this case, however, A_m^* is subject to manipulation and covert control by another agent B^2 who uses implantation processes of some deviant kind. What seems clear about this case is that A_m^* cannot be legitimately *held* responsible by B^2 , since B^2 is in fact covertly controlling A_m^* . If this situation was made transparent to A_m^* or any third party it would be correct to say that the demands and expectations that B^2 is making on A_m^* are ones that B^2 decides will be met or violated. B^2 is, therefore, in no position to criticize, evaluate or react to A_m^* in these circumstances.¹⁴ We might say that since B^2 controls A_m^* 's agency there is insufficient causal distance

between them to sustain the reactive stance. Moral communication and responsiveness presupposes that agents are not related to each other as controller and controllee. When a controller takes up an evaluative/reactive stance toward an agent that he controls there is plainly an element of fraud or self-deception going on. The controller \mathbf{B}^2 can only praise or blame himself for the way in which the agent \mathbf{A}_m^* succeeds or fails to be guided by available reasons.

These limitations do not apply to the relationship between \mathbf{A}_m^* and (non-controlling) \mathbf{B}^1 . \mathbf{B}^1 may be aware that there is some (deterministic) causal story to be told about how \mathbf{A}_m^* acquired the mechanism that she is operating with, but whatever it is (a natural or artificial process) all that matters is that \mathbf{A}_m^* is rationally competent and \mathbf{B}^1 does not control her. The contrast in the relations between these individuals may be illustrated as follows:



It is clear, per hypothesis, that \mathbf{A}_m is responsible to \mathbf{B}^1 , and there is no responsibility compromising relationship between them. Moreover, since \mathbf{A}_m^* possesses the same mechanism as \mathbf{A}_m (i.e. \mathbf{M}), and \mathbf{B}^1 stands in the same (non-manipulative) relation to \mathbf{A}_m^* as he does to \mathbf{A}_m , there is no principled basis for \mathbf{B}^1 treating \mathbf{A}_m but not \mathbf{A}_m^* as responsible. (i.e. since the causal origins of \mathbf{M} do not alter or impair how $\mathbf{A}_m / \mathbf{A}_m^*$ deliberate and act, nor result in any relevant change in the relationship between \mathbf{B}^1 and \mathbf{A}_m^*). When we turn to the situation of \mathbf{B}^2 , however, there is a relevant difference in his relationship with \mathbf{A}_m and \mathbf{A}_m^* . Although \mathbf{B}^2 's situation in relation to \mathbf{A}_m is no different from \mathbf{B}^1 's situation, his relation to \mathbf{A}_m^* is that of controller to controllee. Clearly, then, what is compromised in these circumstances is not the responsibility of \mathbf{A}_m^* as such (since both \mathbf{A}_m and \mathbf{A}_m^* stand in the same relation to \mathbf{B}^1), it is the stance that \mathbf{B}^2

takes toward A_m^* that is compromised by the relationship of manipulation and covert control.

Putting this point in the familiar language of P.F. Strawson, we may say that when the relationship between two individuals is one involving covert control (e.g. through deviant implantation procedures of some kind) then the participant stance on the side of the controller is compromised. The controller is not entitled to take a participant stance in circumstances where he (e.g. B^2) decides when reasons, criticisms etc. succeed or fail to move the agent (e.g. A_m^*). For the controller to retain some commitment to the participant stance in these circumstances would clearly be fraudulent or self-deceptive. However, in the absence of any relationship of this kind (e.g. as with B^1 to A_m^*) the participant stance is not compromised. Granted, therefore, that A_m and A_m^* are identical in respect of their capacity for rational self-control, there is no reason to treat one as responsible and the other as not responsible, unless the stance being taken is compromised by a relation of covert control (e.g. as in the case of B^2 but not B^1). If the hard compatibilist strategy is to succeed then it must, I suggest, draw some relevant distinction along these general lines. A distinction of this kind will enable us to explain why manipulation and covert control is intuitively unsettling, without driving us away from the basic hard compatibilist stance.

The position that I have suggested that compatibilists should take in relation to manipulation examples and circumstances of covert control may be described as “selective hard compatibilism”. Selective hard compatibilism accepts that there is some basis to our intuitive worries arising from circumstances of manipulation and covert control. Unlike soft compatibilism, however, the selective hard compatibilist does not concede that agents in these circumstances are not responsible because of the (deviant or abnormal) “history” involved in the way they acquired their reason-responsive mechanisms. What is compromised in these cases is not the agent’s responsibility, as such, but the legitimacy of the stance of holding an agent responsible on the part of those who covertly control him through the (deviant) implantation process. Assuming, however, that the agent’s capacity for rational self-control is otherwise

unimpaired by the process involved, the stance of those individuals who do not stand in the relation of controller to controllee is not affected or compromised by the agent's history of deviant implantation (i.e. in relation to others). Since the agent has all the time-slice properties and abilities of an agent who is fully responsible given a normal history (i.e. in the absence of manipulation and covert control) there is no relevant basis for refusing to take the participant stance towards an agent of this kind (e.g. A_m^*).

V. God, Walden Two and Frankenstein: Modes of Implantation

Having explained the general principles of selective hard compatibilism, it will be useful to consider a few further examples in order to test our intuitions about such cases. Perhaps the most obvious example – one that has an established place in the history of philosophy - is the theological case involving God as a cosmic covert controller, through the act of divine Creation. On one side, some compatibilists have taken the “hard” view that conditions of (divine) covert control or manipulation do not compromise (human) freedom and responsibility. They deny, therefore, that it is intuitively obvious that if God creates this world, and ordains all human action, then we cannot be held accountable to him or any one else.¹⁵ On the other side, there are compatibilists who are clearly less than comfortable with this position. We find, for example, that Hume, in a well-known passage, considers the implications of his own necessitarian doctrine for Christian theology. In particular, he considers the objection that if the series of causes and effects can be traced back to God, then it follows that God and not humans are responsible for any crimes that that occur.¹⁶ Hume's reply to this objection oscillates between the suggestion that in these circumstances God alone is responsible for all that flows from his act of Creation (since he is their ultimate “author”) and the distinct view that in these circumstances God must share responsibility with humans for any actions that we perform. Hume, in other words, oscillates between hard and soft compatibilist commitments on this issue.¹⁷ The source of Hume's discomfort is that he cannot

concede, consistent with his general compatibilist commitments, that (blind) natural causes of an agent's character and conduct would compromise freedom and responsibility. At the same time, there is something "absurd" about the suggestion that God holds humans accountable (in a future state) for events that he ordains. Clearly the unqualified hard ("bite the bullet") response is not one Hume is willing to accept in this case.¹⁸

Although I doubt that Hume was sincerely troubled by this issue, it should be clear that selective hard compatibilist principles provide a solution to this problem in a way that is consistent with Hume's irreligious intent on this topic. That is to say, the selective hard compatibilist view of this situation is that it is indeed illegitimate and inappropriate for God to hold humans accountable in these circumstances, in so far as God covertly controls us and all we do (as per the Creation hypothesis). On the other hand, this concession does nothing to compromise our basic (hard) compatibilist commitments. More specifically, it does not follow from the fact that God is in no position to hold us accountable that we are not (fully) accountable to our fellow human beings in these circumstances. Since we are not covertly controlled by other human beings it is strictly irrelevant whether our conduct and character is ultimately determined by (blind) Nature or by (a personal) God. Nothing about our current abilities or our qualities of character and conduct is altered or affected either way. Therefore, to us (qua humans) this is not a consideration which fundamentally changes our relation with each other or compromises the participant stance that we take towards each other.

The incompatibilist and soft compatibilist critics may find this theological example less than convincing when we try redescribe it in terms of purely human circumstances and conditions. For example, suppose that instead of God serving as a "global manipulator" we imagine a world like Walden Two, where individuals are "engineered" by the methods of implantation or some related technique adopted by the state for its "utopian" ends.¹⁹ How will the principles of selective hard compatibilism fare in this situation?

In cases like Walden Two, selective hard compatibilism draws the following distinction. In so far as within human society there are covert controllers and those who are controlled by them, the former are not in any position to hold the latter responsible or take a participant stance toward them. Given their relationship, the participant stance is not appropriate, since it is the state controllers who are determining how their subjects deliberate, decide and act. (e.g. through controlled implantation procedures). Any stance of evaluation and criticism is, for them (the controllers), not in order.²⁰ However, on the assumption that the subjects of Walden Two are not related to each in this way (i.e. they play no role in the process of implantation and covert control), and still possess unimpaired powers of rational self-control, no restriction of this kind applies. These individuals have every reason to continue to view themselves as free and responsible agents (as they would if they were created through the processes of blind nature) and to take the participant stance towards each other. In this way, and to this extent, within Walden Two conditions of freedom and responsibility will survive.

As in the theological case of divine Creation, the subjects of Walden Two could wake up one morning and be told, by the relevant authorities, that they are all covertly controlled. While they may well be surprised by this, they have no more reason to suddenly regard themselves as standing in a fundamentally different relations with each other in respect of their status as responsible agents than their theologically conditioned counterparts (or, indeed, than they would if they were informed that they are all determined products of blind, natural processes). In neither case do these agents find that their abilities or qualities have been altered or changed. There is nothing that they were able to do yesterday that they cannot do today. What has changed is that these individuals will no longer view themselves as appropriate targets of moral sentiments in relation to the state authorities who control and condition them – since there is evidently something “absurd” about the authorities criticizing and condemning agents who they are covertly controlling.

The incompatibilist and soft compatibilist may remain unconvinced and argue that the examples chosen continue to obscure the real problems here. In both the divine Creation and Walden Two examples we are presented with circumstances of global manipulation, whereby all agents (i.e. the “normal” or “ordinary” agent) is being covertly controlled. However, if we consider an isolated individual case of covert control, our intuitions may change. Viewed from this perspective, the case of an artificially designed agent who is covertly controlled by his creator is obviously problematic. Let us call cases of this sort “Frankenstein-type examples”, in order to highlight this familiar and “troubling” theme in both literature and film.²¹ Cases of this kind, it may be argued, make clear, that something “abnormal” and “disturbing” is taking place. Surely, our critic continues, no one will deny that Frankenstein-type agents, as described, could be viewed as free and responsible from any point of view? “Biting the bullet” in cases of this kind is an act of philosophical despair - or at least a sign of an inability or unwillingness to think imaginatively about cases of this kind.

The first thing we must do, in order to get clear about the significance of these Frankenstein-type examples, is to eliminate features of the example that are strictly irrelevant or misleading.²² In the first place, it is important to note that the moral qualities of the covert controller may vary – they may be good, evil or mixed. (The same is true in cases of global manipulation, as described above: e.g. either God or the Devil may rule the world.) The case, as described, leaves this issue open. Second, the literature concerned with examples of this kind often suggest not only the shadow of *evil* manipulators operating in the background, they also typically conjure up the image of a “monster” or “freak” who serves as the agent involved (e.g. as in “Frankenstein”). However, cases of this kind are strictly irrelevant since, per hypothesis, we are concerned with individuals who have time-slice properties that are entirely “normal” and present them as otherwise fully functioning and complete agents (i.e. in the absence of any worries about manipulation and covert control). With these distortions removed, we are now in a better position to test our intuitions about such cases and the intuitive force of the principles of selective hard compatibilism.

Clearly selective hard compatibilism does not license any unqualified hard compatibilist approach to these cases. The individual who covertly controls the agent is in no position to take up the participant stance towards this individual – no matter how “complex” or “robust” the capacities and qualities of the (created) agent may be. Having said this, similar constraints and limitations do not apply to other individuals who stand in a relevantly different relation to the agent (“Frankenstein”). Given that the agent is not in any way impaired in his powers of rational self-control (i.e. he is not “abnormal” or “monstrous” in time-slice terms), and he is not covertly controlled by these other individuals, then the agent remains an appropriate target of their reactive attitudes. For these individuals, therefore, the participant stance is not ruled out or compromised simply on the ground that some other individual covertly controls him. From the perspective on other non-controlling individuals, it is immaterial whether the agent’s mechanism M is blindly implanted by Nature or implanted by a covert controller. How the agent functions, and how he relates to those of us who do not covertly control him, is entirely unaffected. Indeed, the case for selective hard compatibilism may be put in stronger terms. Any policy that demands that this agent be treated with a systematic “objective” attitude, simply on the basis of the origins of his (artificially implanted) mechanism, is itself intuitively unfair. Such a policy fails to recognize and acknowledge that the agent as an agent and treats this individual as if he lacks capacities that he clearly possesses (and as such constitutes a form of discrimination).²³

Our critic may persist that the relevant points have still not been covered. Consider, for example, discovering that you are a creation of Dr. Frankenstein and that he has implanted you with some mechanism M and thereby covertly controls you. Even if you do not assume that Dr. Frankenstein is evil, and you accept that this discovery in no way implies that you have suddenly been transformed into a “monster” of some kind (i.e. as judged by time-slice criteria), surely there remains something deeply disturbing about this discovery from the point of view of the agent? More specifically, what we find disturbing about this situation has nothing to do with the stance of the people who may or may not

hold the agent responsible, it has everything to do with how the agent must regard himself.

If there is something about the discovery that we find disturbing, from the agent's point of view, it must be judged in relation to cases where the mechanism M has been blindly implanted by natural processes and there is no possibility of covert control. So the relevant question we should be asking in this situation is what does the agent have to worry or care about in these circumstances? The agent may well find the discovery of a history of artificial implantation and covert control disturbing on the ground that the moral qualities of the covert controller will indeed matter to the way that he functions and operates as an agent. In respect of this issue, the agent may be lucky or unlucky. Naturally, in these circumstances the agent would want to be created and controlled by a benevolent and good creator (just as in the parallel theological situation we would prefer that God and not the Devil is arranging the order of things). In general, if we are being covertly controlled the best we can hope for is a controller who directs our reason-responsive capacities in some desirable way, as judged from our own point of view and that of those other individuals who must deal with us. Notice, however, that an agent who is blindly implanted through natural processes, without any possibility of covert control, will also have a parallel worry about whether he has been lucky or unlucky in the way these (blind) forces of nature shape his character and conduct. He will want to discover what these forces are and he will hope that they are benign and not malevolent in their outcomes (i.e. as judged from his own point of view and that of the people who he is dealing with).²⁴

It is clear, then, that the agent's discovery that he has been implanted and is covertly controlled will raise some distinct issues for him – concerns that do not arise in the “normal” case. However, it is by no means obvious that this discovery must lead the agent to cease viewing himself as a free and responsible agent. On the contrary, if he thinks carefully about his situation he will see that his (time-slice) capacities are not changed or impaired in any way. Moreover, any worries he may have about the (artificial) causal factor that conditions and

directs the process of mechanism acquisition have clear counter-parts in the case of agents who are subject to a process of blind, natural implantation – such as he took himself to be until the moment of discovery. Finally, the agent in question may also recognize that his situation does not necessarily leave him worse off than his “normal” counter-part who is not covertly controlled. It may well be that the “normal” agent is (deeply) unlucky in the natural causes that shape the process of mechanism acquisition, whereas it remains an open question whether the agent who is implanted and covertly controlled is lucky or unlucky. Nothing about the agent’s circumstances, as described, determine this issue and it is this issue that the agent has most reason to care about.

These observations suggest that when Frankenstein-type scenarios are accurately described, they are not disturbing in any way that implies that the agent involved must lose all sense of himself as a free and responsible individual (i.e. not unless he has similar reason for this worry based on blind implantation processes). Faced with these observations, the critic may try one last manoeuvre. If we turn to our perspective as observers in these circumstances – where we are neither controller nor controllee - we may find the principles of selective hard compatibilism are unconvincing. According to this position, the critic continues, conditions of covert control are relevant only to the individual who actually covertly controls the agent. More specifically, this situation is otherwise irrelevant to all third party observers, since they can continue to take up the participant stance unconcerned about the circumstances of covert control. Surely the knowledge that an agent has been artificially implanted by another individual, who covertly controls him on this basis, cannot be dismissed as irrelevant to our attitude and (third party) perspective on this agent? Once this relationship becomes transparent it will matter to us – in our capacity as observers.

Once again, this line of objection is mistaken. For reasons made clear in our earlier replies, it may be important for us to know that an agent is artificially implanted and covertly controlled, given that the moral nature of the controller (e.g. the aims and purposes involved) will indeed matter to us – just as this will

also matter to the agent. The covert control may be benign and benevolent, or it may not. We certainly have reason to care about that. At the same time, however, even if no artificial implantation or covert control is present, we have good reason to care about and investigate the blind, natural causes that may condition and shape the agent's powers of rational self-control. The nature of these causal origins, and the character of the upshots that they bring about, will matter to us either way. It is, therefore, a mistake to think that because we have reason to care (and worry) about whether an agent is being covertly controlled, and what the character and purposes of his covert controller may be, it follows that we cannot view this individual as a responsible agent. No such conclusion is implied, any more than it follows that because we care about the nature of the blind causal forces that may shape and condition an agent's mechanism we must cease to view him as a free and responsible individual. The observer or third party interest in both cases is similar. Whether the process involved is artificial or natural – whether it leaves open the possibility of covert control or not - the agent continues to enjoy the same capacities and abilities either way. Nor is our own relationship with the agent changed or altered in any relevant way by this discovery. We have, therefore, no relevant grounds for abandoning the participant stance towards this individual simply on the ground of the history of artificial implantation and covert control (i.e. unless we also entertain some further sceptical doubts about the implications of any process of implantation, be it blind or covertly controlled). In these circumstances all that has changed is our specific understanding of the particular history involved in the agent's mechanism acquisition – something that is of interest and importance to us (and the agent) whether covert control is involved or not.

The examples that we have considered in this section make clear that selective hard compatibilism involves a distinct set of commitments from those of either soft or hard compatibilism. Unlike hard compatibilism, selective principles acknowledge that circumstances of implantation and covert control have implications that are relevant to issues of responsibility. The way that it interprets these issues is, however, very different from the soft compatibilist

approach. Whereas the soft compatibilist follows the incompatibilist in holding that responsibility in these circumstances is systematically undermined, the selective hard compatibilist maintains that what is compromised is not the agent's responsibility as such, but the legitimacy of the stance of holding the agent responsible on the part of those who covertly control him. Beyond this, however, there is no general failure of responsible agency. It follows from this that compatibilist accounts of rational self-control cannot be said to fail simply on the ground that they permit covert control (i.e. just as orthodox hard compatibilists have maintained). In this way, selective hard compatibilism blocks the fundamental conclusion that incompatibilists are anxious to establish – a conclusion that motivates soft compatibilist attempts to add historical requirements to conditions of responsible agency.

VI. Nothing too Hard to Swallow – Some Final Thoughts

Let us now review the argument of this paper. We began with a discussion of the problems that contemporary compatibilist theories of rational self-control face when presented with cases of implantation and covert control by another agent. One way of trying to deal with this problem is to embrace soft compatibilism and accept that situations of this kind do indeed compromise the agent's freedom and responsibility. The soft compatibilist claims that what this shows is that we must also care about the "history" involved in the way that we have acquired our reason-responsive dispositions. Considerations of this kind, they suggest, will allow us to draw the relevant distinctions we need to make in this area. I have expressed scepticism about this approach on the ground that it remains unclear why, on this account, we should care about the agent's history when it makes no difference to the way that the agent actually deliberates, decides and acts. Having the "right" history does not provide any form of "enhanced freedom", nor will it satisfy incompatibilist worries about having effective control over the process of mechanism acquisition (i.e. given regress problems etc.).²⁵

Granted that the strategy of soft compatibilism based on history fails, we must choose between incompatibilism and some form of hard compatibilism. The hard compatibilist accepts the “implantation standard”: any adequate account of the capacities associated with freedom and moral responsibility must be such that they could be implanted by either natural or artificial processes without compromising the agent’s standing as free and responsible (this being a standard that libertarians claim they are able to satisfy). The obvious difficulty for the hard compatibilist, however, is that this seems to open up the door to covert control and manipulation. I have argued that it will not do for hard compatibilists to *simply* “bite the bullet” on this issue. Something more plausible must be said about the basis of our intuitive discomfort with this situation and why we resist this suggestion.

The strategy I have defended involves drawing a distinction between those who can and cannot legitimately hold an agent responsible in circumstances when the agent is being covertly controlled (e.g. through implantation processes). What is intuitively unacceptable, I maintain, is that an agent should be held responsible or subject to reactive attitudes that come from another agent who is covertly controlling or manipulating him. This places some limits on who is entitled to take up the participant stance in relation to agents who are rational self-controllers but nevertheless subject to covert control.²⁶ In this way, what is compromised by conditions of covert control is not the responsibility of the agent as such. It is, rather, the participant stance of those other agents who covertly control him. Clearly it is possible to establish these specific limits on who can hold these agents responsible without denying that the agents themselves remain free and responsible. When we take this approach we will find that we are no longer faced with an unattractive choice between simply “biting the bullet” or having to “spit it out”. All we need to do is chew carefully, until there is nothing left that we find too hard to swallow.

FOOTNOTES

* I am grateful to audiences at the Fifth European Congress for Analytic Philosophy (Lisbon, 2005); Inland Northwest Philosophy Conference (Pullman, WA., 2006); University of British Columbia, McGill University, and Queen's University at Kingston for their helpful comments and discussion relating to this paper. For further discussion and correspondence I would also like to thank Joe Campbell, John Fischer, Arash Farzam-kia, Ish Haji, Josh Knobe, Rahul Kumar, Storrs McCall, Michael McKenna, Alistair Macleod, Alfred Mele, Jeff Pelletier, Saul Smilansky, and Kip Werking.

(1) See, e.g., Daniel Dennett, *Elbow Room: The Varieties of Free Will Worth Wanting* (Oxford: Clarendon Press, 1984); Susan Wolf, *Freedom Within Reason* (New York & Oxford: Oxford University Press, 1990); R. Jay Wallace, *Responsibility and the Moral Sentiments* (Cambridge, Mass.: Harvard University Press, 1994); John M. Fischer and Mark Ravizza, *Responsibility and Control: A Theory of Moral Responsibility* (Cambridge: Cambridge University Press, 1998). For a critical discussion of these theories (and others) see my "Pessimists, Pollyannas and the New Compatibilism", in Robert Kane, ed., *The Oxford Handbook of Free Will* (New York & Oxford: Oxford University Press, 2002).

2. See Fischer and Ravizza, *Responsibility and Control*, which aims to provide an acceptable answer to this problem.

3. In relation to this problem see Dennett's well-known discussion of the "nefarious neuro-surgeon" (*Elbow Room*, 8). Dennett takes this example from John M. Fischer, "Responsibility and Control", reprinted in J.M. Fischer, ed., *Moral Responsibility* (Ithaca & London: Cornell University Press, 1986).

4. In these circumstances the agent is no longer sensitive or capable of being guided by any consideration other than the implanted desire. For example, a hypnotized agent

may start to act in some arbitrary or random manner when some relevant psychological “trigger” is pulled. This way of dealing with problems of manipulation and behaviour control is discussed in Wallace, *Responsibility and the Moral Sentiments*, 176f and 197f.

5. This terminology comes from Fischer and Ravizza, *Responsibility and Control*, 38: “...although we employ the term ‘mechanism’, we do *not* mean to point to anything over and above the process that leads to the relevant upshot...” In a note to this passage, Fischer and Ravizza go on to say that they “are not committed to any sort of ‘reification’ of the mechanism; that is, we are not envisaging a mechanism as like a mechanical object of any sort. The mechanism leading to an action is, intuitively, the way the action comes about; and, clearly, actions can come about in importantly *different ways*.” (emphasis in original)

6. For a more detailed discussion of this point see my “Critical Notice of John M. Fischer and Mark Ravizza, *Responsibility and Control*”, *Canadian Journal of Philosophy*, 32 (2002), 587-606 (esp. sect. v).

7. Philosophers of quite different views and commitments accept that it is simply intuitively obvious that a manipulated agent cannot be responsible. See, e.g., Robert Kane, *The Significance of Free Will* (New York & Oxford: Oxford University Press, 1996), 65f; Derk Pereboom, *Living Without Free Will* (Cambridge: Cambridge University Press, 2001), 112f; Fischer and Ravizza, *Responsibility and Control*, 194-202, 230-9; Ishtiyaque Haji and Stefaan Cuypers, “Moral Responsibility and the Problem of Manipulation Reconsidered”, *International Journal of Philosophical Studies*, 12 (4), 439-464.

8. Robert Kane, *The Significance of Free Will*, 67-9. This terminology is Kane’s.

9. Fischer and Ravizza, *Responsibility and Control*, Chp. 8. Fischer and Ravizza defend their “historicist” views against a number of critics in “Responsibility and

Manipulation”, *The Journal of Ethics*, 8 (2004), 145-177. For an earlier “historicist” (or “externalist”) account of the conditions of responsible agency see, e.g., John Christman, “Autonomy and Personal History”, *Canadian Journal of Philosophy*, 21 (1991), 1-24. See also Alfred R. Mele, *Autonomous Agents: From Self-Control to Autonomy* (New York & Oxford: Oxford University Press, 1995), esp. Chp. 9.

10. In *Responsibility and Control* Fischer and Ravizza describe this required history in terms of a process of “taking responsibility”. There are, they maintain, three required conditions for this process. The first begins with a child’s moral education, as she comes to see herself “as an agent” (208, 210-1, 238). At this stage the child sees that certain upshots in the world are a result of her choices and actions. When this condition is satisfied the child is then in a position to see herself as “a fair target for the reactive attitudes as a result of how [she] exercises this agency in certain contexts” (211). Finally, Fischer and Ravizza also require that “the cluster of beliefs specified by the first two conditions must be based, in an appropriate way, on the evidence” (238). As I explain, my general doubt about an historicist approach of this kind is that it fails to show that the agent has any (final) control over the process of mechanism acquisition – and to this extent it fails to answer incompatibilist objections. I discuss these issues in more detail in my “Critical Notice of *Responsibility and Control*”.

11. The account of “ownership” that Fischer and Ravizza provide is one that leans on the analogy of Nozick’s (Lockean) historical entitlement conception of justice (*Responsibility and Control*, Chp. 7; cp. Nozick, *Anarchy, State, and Utopia* [New York: Basic Books, 1970], esp. Chp. 7.) The general idea, in both cases, is that ownership of something (e.g. a “mechanism”) depends on the historical process involved (as opposed to “current time-slice” considerations). That is to say, what matters to *ownership* is the *way* something was acquired. The difficulty with this analogy, as it relates to Fischer and Ravizza’s “soft compatibilism”, is that on a Nozickean/Lockean theory of property, individuals may (legitimately) come to own property through processes that do not necessarily involve their own activities or consent (e.g. gift, inheritance etc.). In these circumstances ownership is possible even

though the person concerned did not choose the property or select the mechanism - and so may view what is owned as *imposed* upon him. In general, from the point of view of the analogy with Nozickean/Lockean property theory, mechanism ownership is entirely consistent with a lack of control over acquisition.

12. Kane, *Significance of Free Will*, 67. Among the more recent and valuable defences of “hard compatibilism are the following: Gary Watson, “Soft Libertarianism and Hard Compatibilism”, reprinted in *Agency and Answerability: Selected Essays* (Oxford: Clarendon Press, 2004): and Michael McKenna, “Responsibility and Globally Manipulated Agents”, *Philosophical Topics*, 32 (2004), 169-192. For two other interesting compatibilist counter-arguments to the incompatibilist “global control” examples see Manuel Vargas, “On the Importance of History for Responsible Agency”, *Philosophical Studies*, 87 (2006), 351-382; and also Bernard Berofsky, “Global Control and Freedom”, *Philosophical Studies*, 87 (2006), 419-445.

13. See, e.g., Wallace, *Responsibility and the Moral Sentiments*, 5-6, 15-6, 93-5.

14. In the language of P.F. Strawson, we may say that \mathbf{B}^2 must take an “objective”, not a “reactive” stance to \mathbf{A}_m^* (cp. Strawson, “Freedom and Resentment”, reprinted in G. Watson, ed., *Free Will*, 2nd ed. [Oxford: Oxford University Press, 2003], esp. sect. iv.) Since \mathbf{B}^2 is manipulating \mathbf{A}_m^* by deciding when and which reasons will or will not move \mathbf{A}_m^* , it would be *fraudulent* or *self-deceptive* for \mathbf{B}^2 to adopt a reactive stance. \mathbf{B}^1 is not, however, constrained in these ways.

15. For a classical statement of this (“hard compatibilist”) view see Thomas Hobbes, “Of Liberty and Necessity”, reprinted in *The English Works of Thomas Hobbes*, 11 Vols., W. Molesworth, ed. (London: John Bohn, 1839-45), IV, 248f.

16. Cp. David Hume, *An Enquiry Concerning Human Understanding*, T. Beauchamp, ed. (Oxford: Oxford University Press, 1999), 8.33-6.

17. See my discussion of this point in *Freedom and Moral Sentiment* (New York & Oxford: Oxford University Press, 1995), 160-3.

18. Hume's strategy was to show that this dilemma reveals the *absurdity* of the "religious hypothesis". To this extent his apparent worry about this problem is insincere.

19. B.F. Skinner, *Walden Two* (New York: Macmillan, 1962). See also Kane's illuminating discussion of the significance of *Walden Two* from an incompatibilist perspective: *The Significance of Free Will*, 65-9.

20. A similar situation may be imagined when an author enters into a dialogue with one of his own (fictional) characters. Should the author adopt a critical, reactive stance towards such a character, *we* might suggest to the author: "If you don't like this character, why criticize him, just *change* him." (It is true, of course, that the author may take up the reactive stance in so far as he *pretends*, to himself and others, that he is *not* the creator and controller of this character – then his reactive stance may seem more appropriate.)

21. What I have described as "Frankenstein-type examples" should not, of course, be confused with widely discussed "Frankfurt-type examples", as associated with Harry Frankfurt's influential paper "Alternate Possibilities and Moral Responsibility", reprinted in Watson, ed., *Free Will*. Frankenstein-type scenarios have more in common with cases like the "nefarious neuro-surgeon" referred to in note 3.

22. Cp. Dennett, *Elbow Room*, 7-10, which is especially effective in identifying the (incompatibilist) misuse of "bogeymen" in cases of this general kind.

23. Keep in mind here that, per hypothesis, if the mechanism M were *blindly* implanted by natural processes the agent would be regarded as *fully* responsible and a fit target of reactive attitudes.

24. Related to this, consider how we may be lucky or unlucky in the parents we have, in so far as they may greatly influence the process of mechanism M acquisition. Evidently we *care* about this – and we all hope to have had good parents (and, more generally, we want to be “lucky” in respect of the way we have been brought up).

25. I take this observation to apply not just to the particular historicist approach that Fischer and Ravizza defend, but also to a number of other historicist approaches that I have not directly discussed. This includes, for example, Christman, “Autonomy and Personal History”; Mele, *Autonomous Agents*, esp. Chps. 9 and 10; and Haji and Cuypers, “Moral Responsibility and the Problem of Manipulation Reconsidered”. These “historicist” (or “externalist”) approaches vary in significant and interesting ways, and each deserves consideration in its own right. Nevertheless, my primary concern in this paper is not to provide a series of refutations of these various historicist/externalist strategies but rather to sketch an *alternative* compatibilist approach that avoids the need for any historicist/externalist commitments (and also avoids simply “biting the bullet”).

26. Josh Knobe has suggested to me (in correspondence) that this general conclusion of “selective hard compatibilism” receives some significant support from experimental data. More specifically, the data concerned shows that people who stand in different relationships to the agent have different views about whether or not the agent is morally responsible. One way of reading this is that they just disagree with each other about whether or not the agent really is responsible. However, another way of interpreting the data is that although everyone agrees about the agent’s moral status, they also believe that the agent may be held morally responsible by some but not by others (i.e. depending on their relationship). For more on this see Joshua Knobe and John Davis, “Strawsonian Variations: Folk Morality and the Search for a Unified Theory”, in J. Davis et al eds., *The Oxford Handbook of Moral Psychology* (Oxford: Oxford University Press, forthcoming).

Abstract:*Selective Hard Compatibilism*

Recent work in compatibilism has developed various versions of rational self-control or reasons-responsive theories of moral capacity. One important objection that these theories face is that general capacities of this kind could be, in theory, “implanted” into an agent, thereby allowing covert, non-constraining control. Intuitively, an agent who is being controlled or manipulated in this way is not responsible. “Soft compatibilists” have appealed to the role of “history” and the process of “taking responsibility” as a way of dealing with this general difficulty. In this paper I express scepticism about compatibilist (historicist) approaches of this kind and sketch an alternative “hard compatibilist” approach. I argue that while it is intuitively problematic that an agent is held responsible by another agent who does (covertly) control or manipulate him through an implantation process, there is no similar basis for others who do not control the agent in this way to treat him differently unless there is some occurrent or time-slice incapacity produced by the implantation or covert control process.

