BIROn - Birkbeck Institutional Research Online

FREEDOM AND INDOCTRINATION

MICHAEL GARNETT

It has been alleged that compatibilists are committed to the view that agents act freely and responsibly even when subject to certain forms of radical manipulation. In this paper I identify and elucidate a form of compatibilist freedom, *social autonomy*, that is essential to understanding what is wrong with ordinary indoctrination and argue that it also holds the key to understanding what goes wrong in more fanciful manipulation cases.

I. THE MANIPULATION ARGUMENT

Imagine a causally predetermined action that meets all of the standard compatibilist requirements on freedom and responsibility, such as relevant knowledge, reasons-responsiveness, higher-order endorsement, non-coercion, and so on. Then suppose that the distant causal chains inevitably leading to this action's performance were in fact initiated by some sinister manipulator. Since this further supposition is not precluded by any of the standard compatibilist requirements, the compatibilist looks committed to the implausible view that agents insidiously controlled by unwelcome manipulators may nevertheless be paragons of freedom and responsibility. Moreover, the root source of the compatibilist's problem seems to be her very claim that freedom and responsibility are compatible with determinism. For this is, after all, just the claim that freedom and responsibility are compatible with 'manipulation' by natural forces—and once it is allowed that agents may be free and responsible despite being no more than marionettes, what can it matter who or what is pulling the strings?

This is the incompatibilist's *Manipulation Argument* (Kane 1985, Ch. 3; 1996, Ch. 5; Kapitan 2000; Pereboom 2001, pp. 110-26; Mele 1995, pp. 190-1; 2006, pp. 188-90). Alfred Mele motivates one version of it with a widely-discussed case in which a 'supremely intelligent being', Diana, creates a zygote in a woman, Mary. Diana

> combines [the zygote] *Z*'s atoms as she does because she wants a certain event *E* to occur thirty years later. From her knowledge of the state of the universe just prior to her creating *Z* and the laws of nature of her deterministic universe, she deduces that a zygote with precisely *Z*'s constitution located in Mary will develop into an ideally self-controlled agent [Ernie]

who, in thirty years, will judge, on the basis of rational deliberation, that it is best to *A* and will *A* on the basis of that judgement, thereby bringing about *E*. (2006, p. 188)

Many are reluctant to concede that, when Ernie *A*s, he does so with full freedom and moral responsibility. Yet this is precisely the claim to which the compatibilist seems unavoidably committed.

Although the Manipulation Argument is generally taken to target compatibilist accounts of both freedom and responsibility, to date the focus has mostly been on the latter (indeed, the reference to freedom is frequently redundant, insofar as it is restricted to freedom 'in the sense required for moral responsibility'). Yet the argument also poses a formidable challenge to compatibilist attempts to understand human freedom in its own right. Intuitively, Ernie does not act with full freedom, and yet he may be free in all of the most familiar compatibilist senses. He may enjoy a wide range of alternative possibilities, in the sense that he would have acted differently had he chosen to do so. He may be partially self-created, in the sense that his current character is partly a product of past choices. He may enjoy a substantial degree of practical rationality. He may have a large degree of political liberty. For these reasons, explaining his apparent lack of freedom looks to be every bit as difficult for the compatibilist as explaining his apparent lack of responsibility.

In this paper I take up this challenge, identifying and elucidating a form of compatibilist freedom that agents such as Ernie distinctively lack. On this conception, freedom is essentially a matter of not being *subject to foreign wills*. I argue that agents like Ernie are subject to foreign wills to a degree to which ordinary agents are not, a result that follows from a proper understanding of the distinction between education and indoctrination. I also suggest that, although independence from foreign wills is not itself necessary for moral responsibility, understanding how the Manipulation Argument fails with respect to freedom nevertheless helps to bolster existing compatibilist responses to the argument as regards responsibility.

II. TWO VERSIONS OF THE MANIPULATION ARGUMENT

To fully understand the compatibilist's dialectical position, it is important to see that the Manipulation Argument also causes trouble for certain libertarians (Berofsky 2006, p. 423; Mele 2006, pp. 139-43; Haji and Cuypers 2008, pp. 10-11). For while determinism allows Diana to be absolutely certain that her manipulation will succeed, infallibility is no more a requirement of successful manipulation than it is of any other action. A manipulator

who has made it very likely that an action be performed has, in the ordinary sense, succeeded in engineering or bringing about that action. So suppose, instead, that Diana inhabits an indeterministic universe, and that her knowledge of the current state of the universe and the laws of nature make it possible for her to predict, *with a very high degree of probability*, that *Z* will develop into Ernie and that Ernie will *A* at the appointed time, and suppose that this is indeed what happens. Many will be similarly reluctant to concede that, when Ernie *A*s, he does so with full freedom and moral responsibility.

Given this, it is useful to distinguish between two versions of the Manipulation Argument. On its *narrow version*, it targets only compatibilists; on its *wide version*, it targets both compatibilists and certain kinds of libertarian. Ideally, compatibilists need responses to both versions and a coherent story to tell about the relation between the two. This requires a more detailed analysis of the arguments, a task to which I now turn.

The wide version targets anyone who accepts both of two weak and plausible principles. The first is the following:

> *No Radical Choice:* For a choice or action to be free, and a sound basis for ascriptions of moral responsibility, it must have been somehow influenced by some motivational element.

I use the vague term *motivational element* as a catch-all for such things as preferences, desires, values, and patterns of reasons-responsiveness; that is, for all conative *inputs* to practical deliberation or choice. Note that 'somehow influenced' is much broader than *causally determined*, accommodating causal relations that are merely probabilistic, as well as those that are not event-causal. All that is required of the relation is that it support certain types of counterfactual, namely to the effect that had the agent had different motivational elements then that agent would have acted, or have been more likely to act, differently.

No Radical Choice denies that choice or action undertaken for no reason and motivated by no desire can ever form a sound basis for moral responsibility or be part of a valuable form of freedom. One familiar source of support for this claim lies in the Humean idea that free and responsible actions must in some way spring from one's volitional character, as partially constituted by one's pre-existing motivational elements. Thus a choice that is truly 'radical' in the sense of bearing no connection whatsoever to one's prior character is, far from being an exercise of freedom, no more than a sort of unintelligible

mental spasm, 'a kind of random upsurge of total irrationality' into one's psychological life (Steward 2012, p. 170).

Moreover, *No Radical Choice* entails, by means of a familiar regress, that at least some of our motivational elements are ultimately unchosen. Given this principle, every motivational element is a result of either (i) forces external to the agent's (free and responsible) choices, or (ii) choice made on the basis of motivational elements resulting from (i), or (iii) choice made on the basis of motivational elements resulting from (ii)... and so on. Since every choice must be made on the basis of some pre-existing motivational element, and since no person has made an infinite number of choices—the chain must stop somewhere— what lies at its end must inevitably be motivational elements that are unchosen, having arisen from processes outside of our control.

Nevertheless, this does not yet open the door to the Manipulation Argument, since the relations of influence between the 'given' motivational elements and the agent's choices and actions might be so indirect and weakly probabilistic as to give a potential manipulator no reliable way of governing that agent's behaviour. So views are vulnerable to the Manipulation Argument only if they also endorse a second principle, namely:

> *Strong Influence:* An action may be free, and a sound basis for ascriptions of moral responsibility, even when it has been *very significantly* influenced by its agent's motivational elements.

Suppose that one has absolutely overwhelming reason to act in a particular way: that self-interest, morality, and immediate inclination all converge decisively on the same course of action, such that it is extremely probable, given one's existing motivational elements, that one will act in this way. *Strong Influence* is simply the claim that such an action is not for this reason unfree or one for which one lacks moral responsibility.[1] This is a very plausible claim, since its denial entails that reason itself can undermine freedom and moral responsibility.[2]

Now, if our motivational elements are ultimately 'given', then they might have been 'given' by a manipulator; and if our motivational elements can influence our actions to a very significant extent, then such a manipulator could influence our actions to a very significant

---

[1] This way of putting it passes over some important complexities. For instance, a libertarian might accept that *some* of an agent's free and responsible actions may be strongly influenced by prior states, but deny that *all* of them may be (see, in this context, Kane 1996). I here treat such a stance as a *denial* of *Strong Influence*.

[2] For more on the 'rational cost' of denying *Strong Influence* and its implications for libertarians, see Garnett 2013.

extent. *No Radical Choice* and *Strong Influence* therefore seem to be all that is needed to generate the following, problematic conclusion:

Manipulation: An action may be free, and a sound basis for ascriptions of moral responsibility, even when it has been engineered or brought about by a skilled, sinister manipulator.

Since both *No Radical Choice* and *Strong Influence* are intuitively appealing principles, any view that rejects their conjunction must be somewhat radical. Let me therefore call libertarians who reject one or both principles *radical libertarians*. Hence the wide version of the Manipulation Argument targets both compatibilists and non-radical libertarians.

I turn now to the narrow version. As should be clear, indeterminism is not in itself a solution to the wide version, since its problematic conclusion is not that a manipulator can *deterministically cause* or *make certain* the relevant action, but just that she can reliably bring it about. Nevertheless, all libertarians require some degree of looseness in the connections between actions and prior motivational elements, and this means that manipulators can never be absolutely in control of their victims. By contrast, compatibilists are committed to the view that such control may be ironclad. Compatibilists therefore seem committed to a considerably stronger, and thus more problematic, conclusion, namely:

Manipulation*: An action may be free, and a sound basis for ascriptions of moral responsibility, even when it has been *infallibly* engineered *right down to its tiniest details* by a skilled, sinister manipulator.

So there are two versions of the Manipulation Argument. The wide version is deployable by radical libertarians and by sceptics about freedom and responsibility, and alleges that both compatibilists and non-radical libertarians are committed to *Manipulation*, which is implausible. The narrow version is deployable by libertarians of all stripes as well as by sceptics, and alleges that compatibilists are committed to *Manipulation**, which is highly implausible.

In response, the compatibilist needs to provide a convincing explanation of why *Manipulation** (and *Manipulation*, which it entails) seem implausible. She has two options. One is to explain away *Manipulation**'s implausibility and to convince us that it is true (what

Michael McKenna (2008) calls a *hard-line* response); the other is to reject *Manipulation** and to show that she is not, in fact, committed to it (a *soft-line* response). In addition to this, the compatibilist also faces a further dialectical burden, which is to explain why *Manipulation** seems *less* plausible than *Manipulation*. This further burden, which has gone largely unrecognised, is a particularly tricky one for the compatibilist to lay down, given that the only difference between *Manipulation* and *Manipulation** is the background assumption of determinism.

In the next two sections I leave the issue of moral responsibility to one side and focus just on the issue of freedom. I advance a compatibilist explanation of what worries us about manipulation cases in this regard, this being that there is at least one important, valuable freedom that agents like Ernie lack in virtue of their manipulation. Crucially, I argue that the particular freedom that Ernie lacks turns out to be one that he lacks more of under determinism, despite its being compatible with determinism in general. This enables the compatibilist to explain why *Manipulation** is less plausible than *Manipulation*. Finally, I suggest that the availability of this soft-line response as regards freedom can be used to support a hard-line response as regards moral responsibility.

III. FREEDOM AS SOCIAL AUTONOMY

So in what way does manipulation compromise freedom? A tempting thought may be that manipulated agents lack some kind of *social* freedom, since they are caused to act as they do by other agents. However, we are caused to act by other agents all the time, and are not for that reason unfree. Moreover, we are shaped and moulded throughout childhood by the intentional actions of our carers and educators. When Diana influences Ernie's future actions without his consent, therefore, she does not seem relevantly different from his parents and schoolteachers. It is thus unclear what conception of social freedom could succeed in capturing the fact that Diana undermines Ernie's freedom without also yielding the intolerable result that all carers and educators undermine all of our freedom.

Nevertheless, in this section and the next I advance an account of freedom that avoids this problem. On this account, freedom is essentially a matter of not being subject to foreign wills. This is a form of social, negative self-rule—that is, self-rule as the absence of rule by others—and for this reason is also aptly understood as a kind of *social autonomy*.[3]

---

[3] The following account draws on Garnett 2014a.

Freedom as social autonomy is a *eudaimonic* and not a *deontic* notion. That is, the moral point of the concept is to describe an ingredient in a flourishing human life and not, or not directly, to denote the absence of certain types of moral wrong, or the basis of an agent's possession of certain rights. The idea is that part of what it is for one's life to go well is for one to enjoy a certain kind of independence from the control or manipulation of others. Conversely, to the extent to which one is subject to foreign wills, one is deficient with respect to an important human value. Slaves, inhabitants of brutal dictatorships, victims of domestic abuse, cult members, and the characters of Huxley's *Brave New World* (1932) are all central cases of people whose lives are going badly in virtue of the fact that they are excessively subject to foreign wills.

Importantly, social autonomy is not the only value, and it sometimes competes with other values of equal or greater importance, such as those of community, love, trust, and friendship. Moreover, there is no reason to think that in order to live a flourishing human life one must be *entirely* independent of interpersonal control or be *maximally* socially autonomous. As with many other values (such as knowledge), what primarily matters is that one have *enough*, that one not fall below some threshold such that one's store of the value is inadequate for a minimally decent human life. This is a point to which we shall return.

Being subject to a foreign will is a particular case of being gotten to do something. We may say that *A* gets *B* to *x* just in case *B*'s doing *x* is brought about by *A* acting so as to bring about *B*'s doing *x*. Moreover, there are broadly two ways of getting someone to do something: by modifying a person's options, and by shaping a person's motivational elements.[4] Call these, respectively, *option-oriented* and *preference-oriented* ways of getting people to do things. Importantly, whereas some option-oriented ways of getting people to do things are ways of subjecting them to one's will, as with most threats, others are not, as with most offers. Similarly, whereas some preference-oriented ways of getting people to do things are ways of subjecting them to one's will, as with the kinds of manipulation in which we are currently interested, others are not, as with many forms of rational persuasion. A complete account of interpersonal subjection would provide characterisations of both distinctions. Here I focus only on the latter.

The question, then, is what extra ingredient is needed to turn a mere preference-oriented case of being gotten to do something into a case of being subject to a foreign will.

---

[4] Yaffe (2003, pp. 340-1) makes a similar distinction between interventions that alter the relevant reason-giving facts of the agent's situation, and those that alter the ways in which the agent responds to reasons. For a more fine-grained taxonomy of manipulation cases, see Mandava and Millum 2013.

The crucial idea, I suggest, is that of the other's will being relevantly *foreign*. I propose that the relevant notion of 'foreignness' be understood in terms of the following condition, namely:

> *The conformity of wills condition*: In getting B to x, A does not subject B to A's will if it is the case that (i) B is capable of understanding that A intends to get B to x by certain means and (ii) *were B to know this B* would not repudiate A's intention (where this lack of repudiation is not itself an instance of B's subjection to a foreign will, and where such a recursive application of this condition yields a determinate result).

The basic idea is that if one would endorse the other's will (or, at least, not reject it), then that other will is not relevantly foreign to one's own purposes.[5]

Most rational agents have the general aim of acting in accordance with reasons—indeed, such a general aim may be constitutive of being a rational agent. Thus most rational agents willingly endorse any and all attempts to get them to act in accordance with reasons, that is, attempts at rational persuasion. Conversely, most of us typically repudiate the intentions of those who attempt to influence our behaviour by non-rational means: backroom propagandists, streetwise experts of the 'hard sell', slick virtuosos of the 'soft sell', manipulative pick-up artists, and so on. Nevertheless, there are exceptions. A religious fundamentalist might care more about maintaining her faith than apportioning her beliefs to the evidence, and dismiss all atheist arguments as the devil's work. Destabilising such a person's religious beliefs by forcibly exposing her to such arguments, no matter how rational, would therefore subject her to a foreign will. On the other hand, the will of the hypnotist I hire to help me to quit smoking is not relevantly foreign to me, despite the non-rational nature of her influence. Moreover, many of us are happy for charities to tug on our heartstrings in order to increase our donations (instead of sticking to the facts and figures); and most of us, I would hazard, prefer to be enticed rather than reasoned into bed.

Thus to be subject to a foreign will is to be gotten to do something such that the conformity of wills condition is not met. However, what matters is generally not whether someone is subject to a foreign will *simpliciter*, but *to what extent* someone is subject to a foreign will. In cases in which one is subject to a foreign will, then, the degree of one's subjection is a function of three variables.

---

[5] Note that what must be unrepudiated is not just the intention that B do x, but the intention that B be gotten to x *by certain means*. For more on the intended interpretation of this condition, see Garnett 2014b.

The first is the degree of non-conformity of the wills, that is, the strength of the person's repudiation of the influencer's relevant intention. I am less subject to a foreign will when influenced by another's intention that I only mildly reject than by one that I strongly reject. The second is the extent to which the influence raises the probability of the person's performing the intended action. It is natural to regard the *extent* to which *A* gets *B* to *x* as a matter of the significance of the causal role played by *A*'s action in the etiology of *B*'s doing *x*. Thus if *A*'s action makes only a small difference to the likelihood of *B*'s doing *x*, we may be reluctant to assert that *A* has gotten *B* to *x*; by contrast, if *A*'s action massively raises the likelihood of *B*'s doing *x*, then we are much more inclined to assert that *A* has gotten *B* to *x*. These considerations carry over to cases in which *A* subjects *B* to his foreign will.[6]

The third and final variable is the degree of specificity of the manipulator's intention. The more general the manipulator's intention, the less subject one is to the manipulator's will. This is because the more general the description, the more that is left up to one in terms of how, specifically, one acts under it. To take the limiting case, if you manipulate me just into *performing some action*, then I am subject to your will to only the most minimal degree. By contrast, if you manipulate me into doing something specified right down to the finest detail, then in performing this action I am very much subject to your will. For example, you subject me to your will more in getting me to stay in my room than in getting me to stay in the country.

This is the basic account of freedom as social autonomy. It is in this sense, I shall argue, that Ernie lacks freedom. Yet don't we all lack freedom in this sense?

IV. EDUCATION AND INDOCTRINATION

The problem here is not that all children are subject to unwanted manipulation and therefore turn out to be substantially non-autonomous. This is true, but it is not a problem: children are indeed subject to foreign wills, and insofar as this is in their interests, it is justified paternalism. The problem, instead, is that *adults* may turn out to be substantially

---

[6] According to Yaffe (2003), manipulation reduces freedom because, unlike a natural causal process, it *tracks* the relevant outcome, thus rendering more distant the possible worlds in which the victim has alternative motivational elements. As a result, 'there is a very real sense… in which manipulators limit our options in ways that neutral causal forces do not. And it is that sense in which manipulators take away from our freedom' (2003, p. 344). This is almost right, but the invocation of *options* gets things wrong. Not every possibility is an option: options are objects of choice, and motivational elements are not typically objects of choice. Nevertheless, Yaffe is right that these relative probabilities are important for freedom—just not freedom understood in terms of options. Instead, the fact that manipulation reduces the likelihood of one's having come to have different motivational elements is relevant the extent of one's subjection to foreign wills.

non-autonomous, in virtue of the manipulation to which they were subject as children. When a parent coerces a child—say, into putting her shoes on—he dictates her behaviour in just this instance. But when a parent engineers a motivational element in a child—say, an aversion to throwing food at the table—he dictates her behaviour not just now, but far into the future. So adults, insofar as they are significantly influenced by motivational elements intentionally brought about by their carers and educators, continue to be subject to the wills of those carers and educators. And this means that on the proposed account adults are non-autonomous, even when they are products of the most enlightened and liberal processes of education and socialisation. This looks like an intolerable result.

As we have seen, however, social autonomy comes in degrees, and the question is therefore about which processes of education and socialisation put agents under the threshold of autonomy necessary for flourishing human lives. Indeed, I suggest that one of the main points of the distinction that we intuitively draw between indoctrinating and non-indoctrinating processes of education and socialisation is to separate processes falling on either side of this threshold. As I now demonstrate, the proposed account of social autonomy yields plausible results in this regard.

To see this, consider two hypothetical extremes of educational practice. At *Indoctrination High*, pupils' beliefs and values are instilled as rigid and uncriticisable systems, unshakeably immune to new evidence and acted on without reflection or scrutiny. At *Enlightened College*, propositions are taught provisionally as subject to ongoing processes of criticism and rational revision. The school seeks to develop both its pupils' critical capacities and their motivation to exercise these capacities, so that they will be able to step back from their instilled beliefs, adopt a critical and questioning stance towards them, and, if necessary, go on to revise or abandon them. Call typical adult graduates of these schools *Indoctrinated* and *Enlightened*.

The question is whether the proposed account is able to yield the conclusion that Indoctrinated is substantially less socially autonomous than Enlightened, such that it is plausible to suppose that the former falls under some threshold of social autonomy necessary for living a flourishing human life whereas the latter does not. As we saw above, the proposed account treats the degree of one's subjection to a foreign will as a function of three variables, namely (1) the foreignness of the will, (2) the difference the foreign will makes to the probability of one's acting as intended, and (3) the specificity of the intended act. I now suggest that (2) and (3) hold the key to understanding the relevant differences between Enlightened and Indoctrinated.

I begin with (3), specificity. Teachers at both schools intend their pupils to believe what is true. At Indoctrination High, teachers take themselves to be teaching final, immutable truths (call the set of such truths *P*), and their intention that their pupils believe what is true is therefore extensionally equivalent to the intention that their pupils believe *P*. At Enlightened College, by contrast, teachers take themselves to be teaching provisional 'truths' that are subject to ongoing critical processes of rational reassessment. In their case, the intention that their pupils believe what is true is no more than the intention that their pupils believe *whatever the truth may be*. Moreover, the intention that someone believe *P* is more specific than the intention that someone believe whatever the truth may be. Thus in getting someone to believe the former one exercises more control over her than one does in getting her to believe the latter.

I turn now to (2), probability. In the case of Indoctrinated, instilling an attitude that *p* made it almost certain that, as an adult, he would now act on the basis of *p*, since his closed mind makes revision in light of new evidence unlikely. By contrast, in the case of Enlightened, instilling an attitude that *p* made it only somewhat probable that she would later act on the basis of *p*. This is because Enlightened's more open mind, constituted by her capacity for and motivation to engage in critical reflection, makes it possible that she will at some point revise or abandon this attitude. Teachers at Enlightened College cannot reliably predict what conclusions their students will eventually come to. Since the chains of diachronic control are thereby weaker in the case of Enlightened, she is less subject to the wills of her former educators.

Of course, creating critical thinkers is not the only way of making students' future actions less predictable. A teacher could instead implant a device in their brains that generates beliefs at random. On the proposed account, such students would indeed be significantly independent of their teacher's will, and thus significantly socially autonomous in this respect. Yet they would of course fail to lead flourishing human lives—not due to any deficiency with respect to social autonomy, but due to a crippling lack of rationality. Just as indoctrination is not the only kind of bad teaching, lack of social autonomy is not the only kind of suboptimal life. Autonomy is not the only value.

V. MANIPULATION: NATURAL AND SUPERNATURAL

I return to the Manipulation Argument. As we saw earlier, the compatibilist needs an explanation of why *Manipulation** seems implausible. As regards freedom, we can now see

that it is false and that the compatibilist is not committed to it. This is because manipulated agents are typically deficient in an important compatibilist freedom, namely social autonomy.

In applying this result to agents like Ernie, it is important to understand the differences between ordinary manipulators, such as real-world indoctrinators, and super-powered manipulators, such as Diana. Due to their limited knowledge and means, ordinary manipulators must take diachronic control of others' attitudes by ensuring that those attitudes are relatively static, rigid, and unrevisable. By contrast, super-powered manipulators can do without these crude techniques. This is because their unnatural powers of prediction allow them to assume similar or greater levels of diachronic control while allowing their victims flexible, revisable attitudes—albeit attitudes that will be flexed and revised in predicted ways. Thus rigidity of attitude is just *one* means (despite probably being, in our world, the *only* means) by which an agent can exercise long-term interpersonal control over another. Ernie is therefore every bit as lacking in social autonomy as Indoctrinated, just by different means. For this reason, he has less freedom than he would have had he originated without Diana's involvement.

As we also saw earlier, the compatibilist requires an explanation of what it is about determinism that seems to aggravate whatever troubles us about manipulation cases: that is, an explanation of why *Manipulation** seems even less plausible than *Manipulation*. The idea of social autonomy supplies such an explanation. As we saw, the extent of one's subjection to a foreign will is in part a function of the probability of the manipulator's success. Determinism enables a super-powered manipulator like Diana to engineer Ernie's actions with certainty, something she could otherwise do only with some degree of probability. Ernie is therefore *more* subject to Diana's will in deterministic worlds than in indeterministic ones.

Moreover, this explanation is friendly to compatibilism. It is true that libertarian freewill makes it harder for supernatural beings like Diana to manipulate us, and that were we to live in a world populated by such beings we would have reasons of social autonomy to value its possession. Yet we do not live in such a world, and we are complex and unpredictable enough that the truth or falsity of determinism makes no difference to our vulnerability to ordinary manipulators, who must rely on probabilistic judgements either way. One of the reasons why a focus on super-powered manipulators is rhetorically favourable to incompatibilists is because it directs our attention to contexts in which libertarian freewill has this special value. However, this is not a kind of value possessed by libertarian freewill in our actual context (c.f. Dennett 1984).

In these ways, compatibilists can use the idea of social autonomy to construct a satisfying response to the Manipulation Argument as it applies to freedom. But what of moral responsibility? Moral responsibility does not require social autonomy, so these considerations do not help the compatibilist directly in this respect. Nevertheless, understanding how the Manipulation Argument fails with respect to freedom helps indirectly.

As compatibilists like McKenna (2008; 2014) have argued, the best response to the Manipulation Argument as regards responsibility is, in the end, likely to be a hard-line one. Indeed, various important considerations may be marshalled in support of *Manipulation** with respect to responsibility (see, for instance, Double 1991; Frankfurt 2002; Berofksy 2006; McKenna 2008; 2013; 2014). Yet many compatibilists concede that, even when cast in its best light, *Manipulation** retains a lingering implausibility that such considerations are unlikely ever to fully expunge (McKenna 2014; c.f. Mele 2006, pp. 193-4). What is needed, therefore, is a plausible explanation of how this can be the case if *Manipulation** is actually true (as regards responsibility).

To this end, it is essential to keep separate the following claims:

(1)   People like Ernie may be fair targets for praise and blame.

(2)   People like Ernie may be paragons of agency.

In asserting (1), compatibilists may also be taken to be asserting (2). If so, this is not entirely unfair, given that compatibilists are prima facie just as committed to (2) as to (1). Moreover, there is an easy tendency to move from the idea of one's being *responsible for some action* to that of one's being a *responsible agent*, with the latter then liable to take on the air of a more general ideal of agency (Williams 1995). As we have seen, however, compatibilists have sound reasons for rejecting 2. And once 2 is out of the picture—once it is conceded that *something* is importantly amiss in radical manipulation cases—1 may be an easier pill to swallow. In this way, supplementing a hard-line on responsibility with a soft-line on freedom—a combined strategy we might call a *two-line* response—can perhaps help to bring compatibilists closer to a convincing overall reply to the Manipulation Argument.[7]

Department of Philosophy
Birkbeck College
Malet Street
London WC1E 7HX

REFERENCES

Berofksy, Bernard 2006: 'Global Control and Freedom'. *Philosophical Studies*, 131 (2), pp. 419-45.

Dennett, Daniel C. 1984: *Elbow Room*. Massachusetts Institute of Technology Press.

Double, Richard 1991: *The Non-Reality of Free Will*. Oxford: Oxford University Press.

Frankfurt, H. G. 2002: 'Reply to John Martin Fischer'. In S. Buss and L. Overton (eds.), *Contours of Agency: Essays on Themes from Harry Frankfurt*. Massachusetts Institute of Technology Press.

Garnett, Michael 2013: 'Freedom and Unpredictability'. *Inquiry*, 56 (6), pp. 666-80.

—— 2014a: 'The Autonomous Life: A Pure Social View'. *Australasian Journal of Philosophy*, 92 (1), pp. 143-58.

—— 2014b: 'Autonomy as Social Independence: Reply to Weimer'. *Australasian Journal of Philosophy*, forthcoming.

Haji, Ishtiyaque, and Stefaan E. Cuypers 2008: *Moral Responsibility, Authenticity, and Education*. New York: Routledge.

Huxley, Aldous 1932: *Brave New World*. HarperCollins.

Kane, Robert 1985: *Free Will and Values*. Albany: State University of New York Press.

—— 1996: *The Significance of Free Will*. Oxford: Oxford University Press.

Kapitan, Tomis 2000: 'Autonomy and Manipulated Freedom'. *Philosophical Perspectives*, 14, pp 81-103.

Mandava, Amulya and Joseph Millum 2013: 'Manipulation in the Enrollment of Research Participants'. *Hastings Center Report* 43 (2), pp. 38-47.

McKenna, Michael 2008: 'A Hard-line Reply to Pereboom's Four-case Argument'. *Philosophy and Phenomenological Research*, 77 (1), pp. 142-59.

—— 2013: 'Responsibility and Globally Manipulated Agents'. In P. Russell and O. Deery (eds.), *The Philosophy of Free Will: Essential Readings from the Contemporary Debates*. Oxford: Oxford University Press.

—— 2014: 'Resisting the Manipulation Argument: A Hard-Liner Takes It on the Chin'. *Philosophy and Phenomenological Research*, 89 (2), pp. 467-84.

Mele, Alfred R. 1995: *Autonomous Agents*. New York: Oxford University Press.

—— 2006: *Free Will and Luck*. Oxford: Oxford University Press.

Pereboom, Derk 2001: *Living Without Free Will*. Cambridge: Cambridge University Press.

Steward, Helen 2012: *A Metaphysics for Freedom*. Oxford: Oxford University Press.

Williams, Bernard A. O. 1995: 'Voluntary Acts and Responsible Agents'. In his *Making Sense of Humanity and other philosophical papers 1982-1993*. Cambridge: Cambridge University Press.

Yaffe, Gideon 2003: 'Indoctrination, Coercion and Freedom of Will'. *Philosophy and Phenomenological Research*, 67 (2), pp. 335-56.