

FREEDOM AND MIND CONTROL

David Blumenfeld

I

SOFT determinists from the time of Hume to the present have believed that acting freely is roughly a matter of doing as one wants. Of course, many of these philosophers have disagreed about the precise conditions of free action. Some have claimed that only physical constraint interferes with freedom while others have recognized internal or psychological impediments. Some have analyzed statements about freedom in terms of the basic notion of wanting or desiring, while others have preferred the concepts of willing or choosing, and so on. But, ignoring these matters of detail, there is a core thesis about the relation between freedom and causation that is shared by all of the philosophers in this tradition. The thesis is that one acts freely when one's behavior is caused in an appropriate way by states or events in one's own motivational system. These events themselves may be caused by still other events which lie outside one's motivational system, but that alone does not interfere with our freedom. According to this view, we can act freely even if universal determinism is true and all events are ultimately traceable to causes outside of our control. If the chain of events that leads to the act contains the appropriate motivational links within the agent, then the mere existence of the chain provides no reason for denying the agent's liberty.¹

In this paper I am going to discuss an objection that purports to refute any theory of this type. Many philosophers regard the objection as decisive; and indeed, for a time, it shook my own confidence in compatibilism. I refer to it as the problem of mind control.² This problem is often introduced as a way of substantiating a very old criticism of compatibilist views. As we shall see, however, the old criticism lacks luster until it is allied with considerations about mind control. Then it becomes disturbing enough to call for a careful reply. Here I will first set forth the problem and then offer what seems to me a viable answer.

Let us begin by briefly considering some traditional moves in the free will controversy. Suppose we take as our example of a compatibilist theory the simple view that one acts freely when he does as he wants, or when his behavior is caused by his own desires. It has been objected that satisfying this condition cannot guarantee that an agent acts freely, since it may be that his desires could not have been different. If so, it is said, he could not have acted otherwise and hence was not free. To the reply that the agent could have otherwise if he had wanted to, it has been answered that this still does not settle the question of whether he could have wanted to. To guarantee that one acts freely we would need to specify that the person's desires could have been different, or more specifically, that what the agent wanted was within his own control. But if determinism is true, then none of our desires is ultimately within our own control, since they all stem from factors that existed before we were born.³

As it stands, compatibilists have considered this objection somewhat lame. For they have wondered what sense to make of the requirement that the agent's desires could have been otherwise, or that they were "within his own control." One interpretation of this notion can be spelled out in terms of a counterfactual: we might say that a person could have had other desires (i.e., that his desires were within his own control) if he would have had other desires if he had wanted to have them and therefore had tried to develop them. But the objector will reply that, if the person is really to be free, the desires referred to in the antecedent of any such counterfactual must also be ones whose production was within the agent's control. In this case, however, the objection seems to amount to the requirement that to act freely a person must somehow be the cause of *all* of his desires—the originator, as it were, of his entire character. And this claim has

Prick
5 No 3

very little to recommend it. Accordingly, compatibilists have been inclined to deny that it matters how a person's desires originate. Nor do they think it significant that these desires ultimately derive from causes that are not of one's own making. As long as a person acts on his own desires, and does as he wants, we can legitimately regard him as free.⁴

Despite the forcefulness of this reply, the possibility of mind control throws the whole issue into doubt. To persuade us that the origin of our desires is irrelevant to our freedom the incompatibilist exhorts us to imagine a case in which the source of our desire is *not* in our own control, but instead is in the hands of some other individual—a master hypnotist or an evil genius who has the ability to manipulate our inclinations as he pleases. We may illustrate this graphically by considering what has become a stock example: the controller is equipped with a device that sends waves directly to the agent's brain. By pressing certain buttons he can refashion the person's mental set in any way he likes. Once the right buttons have been pushed, the subject wants, feels, and thinks whatever the controller wishes. He also does whatever the controller wants. For, *ex hypothesi*, the agent's desires cause his behavior and the operator of the device controls these desires. Under these circumstances, the agent does not act freely, since the causes of his desires are outside of his control. But what difference does it make whether the causes of a person's desires are in the hands of a mind controller or merely arise from fortuitous circumstances not prearranged by anyone in particular? Surely it is irrelevant to the question of one's freedom whether one's desires are caused by a conscious agent or by fate. If the causes of our desires are ultimately outside of our control—as determinism implies—then we are never free.⁵

Part of the mind control argument comes out in the following passage from John Wisdom.

Suppose that all your acts are determined by your decisions and your decisions by your knowledge (no doubt imperfect) . . . together with your desires . . . But suppose that on the occasion of each decision the strength of your various desires is fixed by the devil. Suppose that you float a bogus company and ruin thousands. Are you to blame? I believe that you are not. It is a question for inspection. I have confirmed

what I seem to see from inspection by asking others to inspect the same problem. I have carefully asked this question of more than one person who was highly intelligent but sufficiently ignorant of philosophy to have no axe to grind, and have received the reply that in such case you would not be to blame.⁶

In his subsequent remarks Wisdom also implies that it makes no difference to our freedom whether a personal agent fixes the strength of our desires or whether this is determined by impersonal forces outside our control. Many other philosophers agree. For example, after describing a case of mind control similar to the one in the previous paragraph, Richard Taylor says:

Nor does it, of course, matter whether the inner states which allegedly prompt all my "free" activity are evoked in me by another agent or by perfectly impersonal forces. Whether a desire which causes my body to behave in a certain way is inflicted upon me by another person, for instance, or derived from hereditary factors, or indeed from anything at all, matters not the least . . . Wherever it came from, whether from personal or impersonal origins, it was entirely caused or determined, and not within my control. Indeed, if determinism is true, as the theory of soft determinism holds it to be, all those inner states which cause my body to behave in whatever ways it behaves must arise from circumstances that existed before I was born.⁷

Taylor takes it to be evident not only that mind-controlled actions are unfree, but also that the explanation of this is that the ultimate causal origin of these acts is beyond the agent's control. In other words, he holds that mind-controlled actions are unfree *because* they are caused by a chain of events leading to factors that are outside the person's power to affect.⁸ This argument to the best explanation should not be overlooked, since it licenses the inference from the fact that mind-controlled acts are unfree to the unfreedom of any act in a determinist world. Armed with the thesis that mind control undermines freedom *because* it is a cause that is outside the agent's control, the incompatibilist can make a crucial generalization: if mind-controlled behavior is unfree, so is any behavior that is the causal product of factors that are outside the agent's control.

The mind control argument has considerable intuitive force. It does not resolve all of the important questions, since, for example, it gives no explanation of what it means to "have one's desires within his own control." Still, it leaves one with the sense that the compatibilist may be too cavalier in asserting that the ultimate origin of our desires is irrelevant to our freedom.

The argument also appears to be capable of defeating any version of our core thesis that can be constructed. This thesis, we may recall, is that any action caused by appropriate inner states is free, regardless of whether the inner states are brought about by factors beyond the agent's control. But whatever sorts of inner states we decide to say are sufficient for freedom (whether it is the occurrence of certain kinds of intentions, desires, attitudes towards our desires etc.), it seems that the mind controller could cause such states to occur. And if he did cause them, it would be intuitively clear that they are not sufficient for freedom after all. If the argument works, it works across the board, or in a completely general way. So I think we can see the importance of coming to grips with it.

It will be easier to do this if we first state the argument more precisely by setting out its premisses and conclusion. As I formulate it, the argument is about free actions, or actions that are performed freely. For stylistic reasons, in the main text I also use expressions such as "X has liberty," "Y is free," or "Z is in bondage." Whatever other connotations these expressions may have, in this context they refer only to the agent's free, or unfree, actions. (Hereafter "mc" stands for "mind control.")

Argument MC

1. No action that results from mc is performed freely.
2. The best explanation of the truth of 1 is that mind-controlled actions are determined by a causal chain leading back to factors wholly outside the agent's control.
3. If 2 is true, then no action that is determined by a causal chain leading back to factors wholly outside the agent's control is performed freely.
4. Therefore, no action that is determined by a causal chain leading back to factors wholly outside the agent's control is performed freely.
5. Therefore, if all actions are determined by a causal chain leading back to factors wholly outside the agent's control, then no action is performed freely.

This is the argument with which I shall be dealing. It is worth noting, however, that many of its proponents would be inclined to press the issue further by adding:

6. If no action is performed freely, no one is morally responsible for any of his behavior.
7. Therefore, if every action has a causal explanation leading back to factors wholly outside the agent's control, then no one is morally responsible for any of his behavior.

I shall not discuss this addendum, however. My justification is that the argument falters prior to 5. Consequently, we need not probe the issues occasioned by 6.⁹

II

Premiss 1 of this argument depends, as Wisdom puts it, "on inspection," that is, on our having a clear intuition that a mind-controlled agent does not act freely. Although this seems correct at first blush, it remains to be seen whether the case will hold up under more careful inspection.

In fact, the mc example is badly underdescribed. We are told that the agent's desires are caused by waves that affect his brain, but we are not given any detailed information about the character of his mental set at the time at which he acts. Here, as elsewhere, however, details make a big difference. As we shall see, one can fill in the story of the mind controller in a variety of ways, each of which yields a distinct mc situation. The thesis I shall defend is that only in certain of these situations is it plain that the agent does not act freely. In other mc circumstances, it is by no means obvious that he is unfree. Furthermore, all of the cases we examine in which it is obvious that mc undermines freedom can be explained without appealing to the fact that the causes of the agent's desires are outside his control. Thus it is not clear that premiss 1 expresses a general truth: and even if 1 is restricted to uncontroversial cases, it is still doubtful that 2 is true.

Let's see if we can show this, beginning with cases in which it seems evident that mc interferes with freedom.

How, exactly, shall we imagine the mc device affecting the agent's mental set so that it renders

him unfree? One possibility is that the waves generated by the machine cause him to have an irresistible desire to act in a way that he knows is ill-advised and that he would like to avoid. A straight-laced professor is overcome by the impulse to give his lecture standing on his head; an otherwise stolid minister is overwhelmed by the urge to do the Charleston while the choir sings "Just a Closer Walk with Thee." All the work of the mind controller. The impulse is so intense that it dominates the subject's mind and overcomes any reservations he has about the action. However silly, foolish, or improper the act might seem, the agent performs it because he is so strongly impelled to do so. Given the character of the desire, he cannot resist it even if he tries his best. Indeed, we may imagine that the professor and the minister *do* try their best, telling themselves that they will be able to make a laughing stock if they capitulate to the impulse. They try to resist, but to no avail: they are in the grip of an overwhelming and uncontrollable desire. I shall refer to a case of this sort as an "MC-A type case" or, for short, "MC-A."

We may concede that in MC-A the person does not act freely. But we can explain this without bringing in the notion that the ultimate causes of his behavior are beyond his control. Facts about its proximate cause will do. In this example, the agent is moved by a desire that he renounces, but that he is unable to hold in check. Although he rejects the induced impulse (i.e., prefers that it not prevail in action), he simply cannot control it. If we refer to a desire that one rejects as "dystonic," we can say that the person in MC-A is motivated by a dystonic irresistible desire.¹⁰ Now, the presence of such a desire will explain his bondage independently of considerations about how the desire ultimately arose.¹¹ So, there is no need to appeal to the idea that his behavior stems from an unbroken chain of causes leading back to factors outside his control. Thus, although MC-A fits nicely with premiss 1 of Argument MC, it conflicts with premiss 2.

Perhaps we can bring out this point in another way. Consider case MC-A', in which the origin of the desire is within the person's own control. He gets the mc device and, knowing the powers of the machine, presses the buttons that will later give

him an irresistible desire. In other words, he intentionally puts himself in a position like the one forced upon him by the mind controller in MC-A. (Imagine that he is interested in knowing what it would be like to act from this sort of compulsion and presses the buttons for the sake of bringing it on.)¹² Let us also assume that he activates the machine freely, where "freely" is taken in however robust a sense the reader finds it appropriate to supply. Yet surely the assumption that the person freely induces the irresistible impulse does not change the fact that he is unfree when he ultimately acts from it. He was free at an earlier time, but he is not free now, while actually in the grip of the compulsion. We might say that our agent has freely chosen to make himself unfree. This shows that it is the present structure of his mental set that explains the person's lack of freedom in this type of case. It cannot be maintained that the absence of freedom is due to the uncontrollable origin of the desire, since we have imagined that the agent actually did control the origin of the desire.

Since compatibilists characterize free action as behavior determined *in an appropriate way* by events in one's own mental set, they can cover MC-A and MC-A' by stipulating that a dystonic irresistible precludes free action. This restriction, moreover, is an entirely natural extension of the simple view that freedom is the ability to do as one wants. In the cases now under discussion, the agent refuses to identify with the desire that causes his behavior, and he experiences it as something he wishes to oppose. In view of this, he can say, quite ingenuously, that he "really wants" to behave differently—expressing thereby his sense of his own bondage.¹³ It is for reasons such as these that many compatibilists have excluded behavior motivated by dystonic irresistible desires from the category of free action.¹⁴

If the mc situation is thought of along the lines of MC-A, it does no damage to compatibilism. So, let us look for other ways in which the mind controller can operate.

III

It will be instructive to begin with a case that appears to be an alternative *modus operandi*, but

in fact is not adequately described. Supplying the details required to secure its coherence will produce genuine cases and will also guide us toward a suitable compatibilist response to them.

In the putative case, MC-B, the mind controller induces a strong desire to perform an action that the agent would not otherwise dream of undertaking. This desire, however, is not as strong as its counterpart in MC-A: the person would succeed in suppressing it if he thought he had adequate reasons and if he tried his best. But the controller also implants a conviction that the act is acceptable and thus sees to it that the agent will not try to suppress the induced desire. The person will later be surprised by his behavior; he will wonder how he could have gotten such an impulse, or how, having gotten it, he could have failed to suppress it. But at the moment of action he perceives the induced desire as syntonetic and therefore sees no reason to suppress it. A well-known campaigner against homosexuality gives a speech favoring Gay Liberation; a communist hands out leaflets preaching the doctrines of Milton Friedman. Each of them is later astonished at himself. The mind-controller strikes again!

What is the problem with this example?¹⁵ To begin with, we said that the agent will later be surprised that he ever should have had such a desire, much less that he should have acted on it. But, in fact, we should expect him to be surprised by the impulse from its very inception. Let's reconsider the individual who gets the urge to support the views of Milton Friedman. Surely the person will find this inclination bizzare. After all, he's a communist! Everything he stands for commits him to rejecting the impulse. He has what he regards as conclusive reasons for supposing that Friedman's views are not only false, but pernicious. Now we surely can imagine that the controller suddenly gives the agent a feeling that "this Friedman stuff isn't so bad," though it would be misleading, I think, to describe this as a real conviction. Whatever we call the tendency, however, it is one the agent will find in profound conflict with his character and with the whole structure of his other relevant beliefs. It therefore makes no sense to suggest that he will perceive either the impulse or the "conviction" as syntonetic. It is much likelier that he will judge them to be what they in fact are—anomalous

and schizophrenic intrusions into the order of his mental economy. Given this judgment, he can be expected to bring all his powers to bear to refrain from acting on the impulse. But, according to one part of our story, he would succeed in suppressing the impulse if he tried. This implies, contrary to another part of our story, that he foils the mind controller.

Obviously, it will not help to stipulate that he is unable to control the impulse. In that case he acts from a dystonic irresistible desire, and we are back to MC-A.

The principle difficulty in our description of MC-B is that the controller has induced an anomalous impulse, while leaving intact the rest of the individual's dispositions and mental powers. Under these conditions, no account is available of why the agent would not suppress the impulse. We can supply this account, however, and render the case intelligible, by imagining that the controller makes more extensive intrusions into the person's psychology.

More than one strategy for doing this is available. In our next example, MC-B' the controller induces the equivalent of a post-hypnotic suggestion. In the post-hypnotic state, the agent "feels like" passing out capitalist literature. If he recalls, or if he is reminded by others, that he is known for his anti-capitalist sentiments, he offers feeble rationalizations for his behavior—"Oh, that was just a phase" or "Yes, I used to accept that line, but I'm a capitalist now" or the like. These responses have nothing to do with his real motivation, however, nor do they represent any genuine reflection or evaluation on his part. The post-hypnotic suggestion causes his behavior and, in virtue of that suggestion the agent is impervious to reason. Although he may produce a thin veneer of verbal "defenses" for his actions, he is, in truth, unable to assess his impulse or even to entertain reasons for behaving in some other way.

Although MC-B' is perfectly coherent, the compatibilist's reply is obvious. The person is definitely unfree, but the explanation is simply that he lacks the capacity to evaluate his desires and to consider grounds for alternative courses of action. Compatibilists have long since stressed that a certain level of rationality, including the ability to weigh

and consider one's desires, is a condition of human freedom.¹⁶ Thus, in MC-B', the crippling effect of the post-hypnotic suggestion explains the absence of freedom, not the fact that there is a chain of causes leading back to factors outside the agent's control. Indeed, the lesson of MC-A' can be re-applied: the agent would be no freer at the time of action if he had induced the suggestion himself. Premiss 2 fails again.

MC-A and MC-B' have much in common. So much, in fact, that I suspect some people will wonder whether any purpose is served in distinguishing the two cases. For instance, both involve impulses that are contrary to the overall tenor of the agent's personality. It is true that in MC-A the impulse is dystonic, whereas in MC-B' it is not. But, in the latter, the person fails to renounce the impulse only because the controller has blocked his conscious access to his normal attitudes. In fact, the induced desire is out of harmony with the rest of the person's character and it *would* be dystonic if he were in touch with himself.

Furthermore, we can classify the desire in MC-B' as irresistible. Here again, it is possible to draw a formal distinction between the two cases. In MC-B', but not in MC-A, the agent would be able to suppress the impulse if he thought he had adequate reasons for doing so. But, since the controller has rendered him incapable of considering reasons, the difference is not significant *vis-à-vis* the question of resistibility.

Despite their shared features, however, the two examples bring out distinct points. In MC-B' the agent has lost his capacity to assess his desires and thereby is deprived of freedom. In MC-A, the agent retains this ability, but is still unfree because he is under a compulsion to act on an impulse he opposes. MC-B' shows that the capacity to evaluate one's desires is necessary for human freedom; whereas, MC-A shows that this capacity is not sufficient. We therefore do well to keep the cases separate.

The controller's tactics in MC-B' are heavy-handed. He is capable of more artful work. In MC-B' he leaves the agent's power to think and reflect intact and tampers instead with his established tendency to give weight to particular kinds of evaluative considerations. Pursuing our example

of the communist, we might imagine this scenario. (In what follows, we do not aspire to serious political analysis, but only to a quick and easy sketch of the controller's technique). Here is a rough account of the relevant part of the man's mental set before the controller comes onto the scene. The communist acknowledges that the capitalist scheme has a few advantages over his own. He is prepared to concede, for example, that a liberal capitalist democracy affords greater opportunity for individual autonomy and self-expression than does a more tightly controlled communist system. But his considered opinion is that these advantages are not nearly so great as they are advertised to be by his ideological rivals. In his view, the virtues of capitalism are relatively insignificant when compared, for example, to the more equal distribution of goods in a communist society, the lower level of risk and insecurity to the general populace, and so on. As he sees the trade-off of values, communism is clearly preferable.

Now let us imagine that the controller steps in and temporarily alters the way the person views this trade-off. Perhaps he strengthens the individual's sense of the importance of self-expression and personal autonomy, just as he partially suppresses his repugnance to the problems of capitalism. The controller can effect this, I believe, without altering the person's perception of the relevant empirical facts (for example, that capitalism provides more individual autonomy, that communism distributes goods more equally, and so on.) and without necessarily foisting any false beliefs on him. He merely shifts the significance the individual is inclined to give to these facts. There is nothing incoherent in this idea, for the values that are in competition are not obviously commensurable and a preference for one set over the other does not follow automatically from an acceptance of the factual descriptions we have postulated above.

If the controller does this, the agent will experience a change of heart, an apparent conversion, which will last for as long as the manipulator keeps the appropriate buttons depressed. Of course, we have imagined the "conversion" coming about abruptly, and so the agent may be surprised and unsettled by it. If he is cautious, he may be disinclined

to act immediately on his newfound point of view, preferring instead to think things over for a while. But we should not be surprised to see him eventually passing out the Friedmanite literature.

It is significant to note that in MC-B", the person really can "think things over." He can delay action and review the whole communism/capitalism issue, debating with himself, or others, the crucial questions of ideology. It is just that, when he now thinks things over, the net gain in autonomy and self-expression under capitalism will strike him as ultimately worth the loss in other social benefits obtained under communism.

The agent is amenable to reason in still another sense: further factual information of an appropriate kind could change his mind about ideological issues. His current penchant for capitalism is based partly on his newly-enhanced feeling for the importance of autonomy and partly on his assessment of the kinds and the amount of value that would be sacrificed under capitalism. If he found proof that there was a great deal more to be lost under capitalism than he had recognized—if, for example, he discovered that the degree of suffering is twice what he had thought—he would again change his mind. Thus, unlike his post-hypnotically suggested counterpart in MC-B', this individual's opinions and behavior can be influenced by evidence. The controller, however, may know that no suitable evidence is in the offing and thus have nothing to fear from the agent's exercise of reason.

Despite the fact that the agent's effective desires are syntonic, that he is amenable to reason, and so on, I am still inclined to deny that he acts freely. At any rate, I am so inclined when I consider that the controller has induced a temporary "conversion," which he achieves in part by rendering inoperative the dispositions of the person's already formed character. Insofar as we imagine, as we have, that the agent's older and more permanent dispositions persist behind the scenes, and are soon to reassert themselves, I think we must accept the incompatibilist's judgment that the individual in MC-B" does not act freely.¹⁷

Once again, however, the compatibilist is in a position to deny premiss 2 of Argument MC and to account for the absence of freedom in terms of his own core thesis. Acting freely requires that

one's behavior be caused in an appropriate way by one's desires and attitudes. In explaining what they mean by "an appropriate way," however, many compatibilists have stressed that free action is determined by the agent's character. I believe that this idea can be used to deal with MC-B".

Unfortunately, the notion of being determined by the agent's character has not always been made as clear as it needs to be, and it is likely to be misunderstood. It suggests, on the one hand, that a free act must be "in character" and, on the other hand, that it must be caused by factors that are entirely innate or within the person's own makeup. Neither of these interpretations is part of the notion as I would suggest that it be employed.

An act is "in character" if it is the sort of thing one typically does in certain kinds of circumstances. But an act can be determined by a person's character even if the behavior is not typical of him. What is required for the application of this concept is not that the act be in character but that it be the causal product of the various desires, attitudes, and values that have developed in the course of the agent's ongoing history and that form his present makeup. Clearly, a piece of behavior can be the causal product of such factors without being the sort of thing the individual usually does.

Similarly, in saying that a free act is caused by one's character, I do not mean to suggest that it is determined by factors that are exclusively innate, or that external stimuli do not play an important role in bringing the act about. Naturally, our actions are determined in part by the various stimuli that impinge upon us—by the things, for example, that we see and hear. When an individual does something it is obviously no bar to his freedom that he behaves as he does because of what he has noticed, observed, or otherwise taken in. Thus, if I accept your offer of a bribe, it is not an argument against my freedom to point out merely that an external factor (the bribe) played a role in causing my behavior. But if I am to act freely, these external factors must have their effect by working *through* my existing values and desires, so that the action expresses my character (as it obviously does in the case of the bribe). A necessary condition of acting freely, then, is that if external factors cause an agent's behavior, they do so by working through

that agent's character. Perhaps it would be slightly more perspicuous if we were to drop the phrase "caused by one's character" and say instead that free action must be "mediated by one's character."

Having set out this requirement, we can see why MC-B" is not a threat to compatibilism: the agent's act was not mediated by his character. It is true that his behavior was caused by one of his desires, namely, the one the mind controller gave him. It is also true that the controller caused this desire to be syntonetic, that is, one the agent identified with, or affirmed, at the time of action. But these reactions (both the first-order desire and its endorsement) were directly "planted" in the agent's mental set. They did not arise from the constellation of factors comprising the person's already developed character. In fact, the elements of the agent's mental set which did form part of his ongoing character played virtually no role in determining his action, since, by hypothesis, they were temporarily rendered inoperative. What we have, then, is a case of causal determination that by-passes, or short-circuits, the agent's character. Our sense that he did not act freely is grounded in this fact, not in the idea that any behavior is unfree that is traceable to causes outside one's control. Consequently, the compatibilist can deny premiss 2 and argue that the best explanation of the absence of freedom in MC-B" is that the person's behavior is the product of an inappropriate causal chain.

We can reinforce this conclusion by constructing an example that is in some respects like MC-B" but does not involve the hiatus between the proximate cause of the agent's behavior and his previously operative character. Consider case MC-B'" in which someone *persuades* the agent to act in the manner described in MC-B".¹⁸ Instead of using an electrical device to manipulate his subject's brain, the interlocutor simply talks to the agent, presenting arguments that move him and that alter his views. Perhaps he is able to point out things about personal autonomy—for example, ways in which it enriches a person's life and those of others—that the agent had never been able to see before. Perhaps he reveals other positive aspects of capitalism, or negative features of communism, that had escaped the agent's notice or eluded his powers of imagination. At the end of a series of discussions the person's

opinions have changed, and he marches off, clutching his copy of *Free to Choose* in one hand and his pamphlets in the other.

It seems to me that in MC-B'" the agent acts freely, whereas in MC-B" he does not. In both cases, one person causes the behavior of another. The difference, however, is that only in the former case does the process of causation work through the agent's character.¹⁹ This again brings out that it is the absence of mediation that explains the lack of freedom in MC-B".²⁰

It is important to bear in mind that mediation by one's character is only a necessary condition of free action. The condition is that one's established character must be a part of the chain of causes that contributes to the action, or in other words, that one does not act freely if the causal chain leading to action by-passes one's character. The agent's behavior in MC-B" fails to meet this necessary condition because the controller entirely short-circuits the person's character.²¹

This raises the question of how complete the character by-pass must be in order to undercut freedom. It would be easy to imagine mc cases where the short-circuit is not quite so thorough as it is in MC-B", yet where the role of the agent's character is too insubstantial to yield the intuition that he acts freely. It would therefore certainly be wrong to suggest that only a total by-pass precludes freedom. (Indeed, it might be questioned whether the short-circuit in MC-B" itself is 100% complete. Possibly, I have idealized slightly even here.) We are dealing with a continuum of cases in which the absence of freedom is most evident where large and significant elements of one's character have been by-passed. But it will be difficult, if not impossible, to draw a sharp line. There is room for a wealth of intermediate and borderline cases, in some of which our sense may be that there is partial freedom and in others of which people's intuitions may be unclear, wavering, or conflicting. (Imagine that someone talks me into doing something. He convinces me with arguments that appeal to my existing desires, but the talk is very fast and slick. If I had been a little more experienced, or if he hadn't been so smooth, I probably would have thought otherwise of his proposal. Or what about this? Seated in the movies—and not noticing the

mind controller in the wings—I suddenly get the desire to eat a large bag of popcorn. Having no objection to doing so, I decide to gratify the urge forthwith. But I might not have done so had I known that the desire was surreptitiously induced. I resent being toyed with like that.)

Obviously, a great deal would have to be said to sort out the multifarious array of intermediate possibilities. But, fortunately, it is not incumbent upon us to do so here. If a given type of partial mediation turns out to be consistent with freedom, well and good. The incompatibilist cannot use it to support premiss 1 of Argument MC. If a second type of partial mediation undercuts freedom, wholly or in part, then this, too, is well and good. The compatibilist can explain whatever degree of bondage there is in terms of the absence of the appropriate type of mediation. And, if a third type of partial mediation gives rise to unclear, wavering, or conflicting intuitions about freedom, this, again, is no comfort to the incompatibilist, who requires clear-cut examples on which to found premiss 1.

IV

It may seem that our response to the variations on MC-B leads directly into a trap by suggesting a final, powerful strategy for the proponent of the Argument MC. We disposed of the aforementioned cases by arguing that they did not involve the proper mediation by the agent's character. But what is there to prevent the controller from manipulating the agent's character itself? If he does so skillfully, then our position precludes any compatibilist explanation of the person's lack of freedom. For the controller can refashion the agent's character so that it produces action via whatever inner states the compatibilist deems appropriate for freedom.

So let us construct a case along these lines. In MC-C the controller does not manipulate the agent in any of the ways we have hitherto described. Instead, he permanently alters the agent's character, shifting his mental set and changing his psychological makeup in suitable ways. By tampering with the subject's personality, the controller sets things up so that the same sorts of desires arise as the ones that determined the person's will in MC-B". But this time the desires are continuous

with the agent's character and indeed arise *from* it. They are not mere adventitious implants, effective only because of a temporary immobilization of the person's more stable, "normal" personality. We may imagine that the manipulator achieves this through a single act of mind control or that he undertakes it piecemeal, so that the agent undergoes a gradual evolution. But however the manipulator operates, slowly or at a fell swoop, there will be some point at which the agent's character has been changed. At that point, the controller can step out of the picture, assured that the behavior he desires will spring naturally from the person's altered character. If the job is done deftly, the internal states leading to action will involve no impediments to freedom, either as defined here or as might be further elaborated in compatibilist terms. Since his character is ultimately the work of the mind controller, the person's actions are not performed freely. But the compatibilist is in no position to explain this.

The answer is that no explanation is required. When we reflect on MC-C, I believe that we will find no decisive support for the claim that the agent is unfree. It is true that before the buttons were pressed, the person would not have chosen to do the things he now does, for they would have been completely out of line with his character, that is, with his desires, values, and reflective commitments. Therefore, one might be inclined to infer that the agent is not at present acting freely. But the inference is unwarranted, since he no longer *has* his old character. The controller has obliterated it and replaced it with a new one. In fact, in a case like that of our communist-turned-capitalist, it is almost as though the controller has produced a new person. (The metaphor of a new person is appropriate here because the changes are deep and enduring and have ramifications that extend virtually throughout the individual's character. In cases of less profound and extensive changes, the metaphor will be correspondingly less suitable; but this will not affect the issues before us.)

The question whether this "new person" acts freely, however, cannot be answered in the negative merely by observing that someone else has caused his character. Parents, for example, commonly make efforts to mold the character of their children from the very earliest stages of infancy. And some-

times, it seems, they succeed. Yet it is surely implausible to suppose that, to the extent that they succeed, they necessarily deprive their offspring of the ability to act freely. In fact, some of the most commendable parental efforts are aimed at increasing autonomy by producing individuals who are devoid of psychological impediments to freedom. We may, of course, dispute the effectiveness of any particular technique that parents employ. But can it be maintained on purely conceptual grounds that no such techniques could conceivably promote autonomy, since they involve the causation of character traits by another person? On the contrary, to judge whether the person's freedom is ultimately diminished or augmented, we would need to know what *sorts* of effects the parent has had on the child.²² The same can be said of the mind controller: whether he produces a person who acts freely all depends on the kind of psychological makeup with which he endows him.

We will get a better grip on this issue if we picture several individuals who have been endowed by their controller with distinct psychological profiles. In sketching these profiles, I draw on the suggestions of a number of recent writers. But, at root, my outline of a free agent is a very traditional one, which is close in spirit to Plato's depiction of the just person.²³

There are certainly ways of refashioning a person's character by which the controller could produce an individual who does not act freely. Some of these have already been covered, as for example, if he were to produce someone whose psyche is riddled with irresistible dystonic desires or whose character is governed by impulses the person is entirely incapable of assessing rationally. To the extent that one's behavior is so motivated, one's freedom is correspondingly impoverished. Perhaps the same can be said if the controller has wrought a person whose actions are dominated by unconscious impulses which the individual can neither affirm nor reject, modify nor control. The controller might also compromise the agent's freedom by giving him the motivational system of an extreme compulsive who performs certain kinds of acts as the only way of mastering a powerful sense of anxiety. Or, once again, he might restrict his subject's freedom if he were to produce someone who

is motivated like Alex in Anthony Burgess's *A Clockwork Orange* or like Winston Smith at the end of *1984*.²⁴

Now the words "mind controller" have a somewhat ghoulish ring. So, in discussing this issue, we may unwittingly think only of grisly cases like the ones mentioned above. But, of course, a controller need not produce anything remotely like these broken and crippled personalities. By hypothesis, he can alter a person's mental set in any way he pleases.

Therefore, let us put aside such cases and imagine instead as favorable a profile as we can. Let us suppose, in fact, that the controller creates a person who is in no way scarred or psychologically debilitated. He is not subject to any irresistible desires and is not driven either by incapacitating anxiety or by powerful, deeply repressed motives. More positively, we may imagine that his effective desires are syntonetic, or ones by which he is pleased to be motivated. Since he entirely identifies with the desires that move him to action, he has a deep inner sense of his authorship of his behavior and of his own responsibility for it. This sense is amplified by his awareness that his personality and behavior are flexible, and that he has the ability to change his responses to fit the demands of changing circumstances. His capacity to exercise reason is also unfettered. When faced with difficult choices, he is able to consider his alternatives, assess them in the light of all his desires and values, and guide his behavior accordingly. To a modest degree, he is even able to alter his values and personality. If, on occasion, he should become dissatisfied with particular features of his character, he possesses the will to change and to develop new patterns of behavior. Of course, all of this, including his disposition to modify himself under specific kinds of circumstances, is the result of his basic character, which was intentionally given to him by the controller. If he is apprised of this, however, he rejoices in the fact and would not have it otherwise. For he believes, correctly, that his basic character serves him very well.

Perhaps there are further characteristics we could add and, if I have neglected any, other compatibilists may enhance this profile as they see fit. But my own strong intuition is that we have already built in enough to anchor firmly the inner conditions

required for freedom. To be sure, external forces, whether in the form of physical constraint, coercion, or the like may frustrate the agent and prevent him from doing as he wants. But, barring such interferences, I am inclined to think not only that such a person acts freely but also that his freedom is as complete as it could be.²⁵

I shall not defend this last claim, however. The present argument does not require it. Instead, I limit myself to asserting something much more modest, which I think is manifestly true, to wit: that it is by no means obvious that the sort of person we have described is incapable of acting freely. It is important to see that this is *all* that is needed to topple case MC-C. Recall that the basic strategy of Argument MC is to support incompatibilism by first producing a type of example that anyone can see in-

volves an unfree person and then arguing that the best explanation of why he is unfree implies the truth of incompatibilism. If it isn't obvious that the person is unfree, the first premiss of the argument is inadequate.

It seems to me, then, that none of the preceding ways of filling in the details of the mc situation supports Argument MC. Although intermediate and mixed cases abound, I can think of no examples that are superior to the ones I have considered. I therefore attribute the initial plausibility of the argument to a tendency of its proponents to underdescribe mc cases and hence to blur the distinctions drawn here. If this is correct, the problem of mc is not the serious threat to compatibilism that many have believed it to be.²⁶

Georgia State University

Received October 20, 1987

NOTES

1. A few references to classical compatibilist accounts are: John Locke, *An Essay Concerning Human Understanding* (New York, 1959), Vol. I, Book II, chapter XXI; David Hume, *A Treatise of Human Nature*, Book II, part III, section 2 and *An Enquiry Concerning Human Understanding*, section VIII; G. E. Moore, *Ethics* (London, 1912), chapter VI; R. E. Hobart, "Free Will as Involving Determinism and Inconceivable Without It," *Mind*, vol. 43 (1934), pp. 1-27; J. S. Mill, *An Examination of Sir William Hamilton's Philosophy* (New York, 1939), pp. 143-56. The material cited in the last three references is reprinted in B. Berofsky (ed.), *Free Will and Determinism* (New York, 1966), pp. 63-95, 159-73, and 54-64 respectively.

2. The objection has recently come under considerable fire from Daniel Dennett in his penetrating *Elbow Room* (Cambridge, Massachusetts, 1984). See also his *Brainstorms* (Cambridge, Mass; 1981) especially pp. 246 f. I am very much in sympathy with Dennett's efforts to disarm the mind control gambit. But our approaches on this topic differ in a crucial respect: Dennett does not regard the gambit as an *argument*; he treats it instead as a member of a family of thought experiments which he refers to as "intuition pumps." Speaking of intuition pumps, he says: "Such thought experiments (unlike Galileo's or Einstein's, for instance) are *not* supposed to clothe strict arguments that prove conclusions from premises. Rather their point is to entrain a family of imaginative reflections in the reader that ultimately yields not a formal conclusion but a dictate of 'intuition'." (*Elbow Room*, p. 12. Italics in original.) Dennett's strategy, then, is not to reconstruct and refute a mind control argument, but simply to adduce considerations that will derail the reflections entrained by a seductive "thought experiment." For three reasons, I proceed in a more traditional way. First, unlike Dennett, I do not know how to get a handle on an intuition pump. Second, whenever someone recommends a position to me by providing considerations on its behalf, I regard it as fair to ask: "Now what's the argument exactly?" If none is forthcoming, or if I am told that there really isn't any argument, I feel entitled to demur. So, whatever the intentions of those who operate the pump, it may be useful to make of it the best, or most plausible, argument we can. Finally, Dennett's claims to the contrary notwithstanding, some proponents of the mind control idea certainly have treated it as an argument. The most interesting and developed example I can point to is found in Robert Kane's provocative *Free Will and Values* (Albany, 1985). See especially chapters III and IX. *Free Will and Values* was published after *Elbow Room*, but see also John Wisdom, *Problems of Mind and Matter* (Cambridge, 1934), chapter VIII and Richard Taylor, *Metaphysics* (Englewood Cliffs, 1974), chapter V.

3. Taylor, *op. cit.*, p. 51.

4. Hobart, *op. cit.*, especially pp. 83-84 in *Free Will and Determinism*.

5. Taylor, *op. cit.*, p. 51.

6. *Op. cit.*, p. 116. In the passage quoted here Wisdom mentions only the issue of blame, but the context indicates that he thinks the Devil's action undercuts blame because it seriously compromises the agent's freedom.

7. *Op. cit.*, p. 51. Pretty much the same idea is echoed by Robert Kane when he has an imaginary character (who speaks for the author) say to a victim of mind control: "... control by a controller made you unfree because it was determining ... If now we tell you that all along there was no controller, but the determining circumstances were created by natural forces, you are not to infer that you were more free all along." (*Op. cit.*, p. 43) (Italics are in the original.)

8. Cf. Wisdom, *op. cit.*, p. 118.

9. There is a voluminous recent literature on the question "In what sense, if any, does moral responsibility presuppose freedom?" The following is a small sample of articles from this literature: Daniel Dennett, *Elbow Room* (*op. cit.*); Harry G. Frankfurt, "Alternate Possibilities and Moral Responsibility," *Journal of Philosophy*, vol. 66 (1969), pp. 829-39; "Freedom of the Will and the Concept of a Person," *Journal of Philosophy*, vol. 68 (1971), pp. 5-20, and "Three Concepts of Free Action II," *Proceedings of the Aristotelian Society*, supplementary volume (1975), pp. 239-51; David Blumenfeld, "The Principle of Alternate Possibilities," *Journal of Philosophy*, vol. 68 (1971), pp. 339-45; Gerald Dworkin, "Acting Freely," *Noûs*, vol. 4 (1970), pp. 367-83; Don Locke, "Three Concepts of Free Action I," *Proceedings of the Aristotelian Society*, supplementary volume (1975), pp. 95-112; P. S. Greenspan, "Behavior Control and Freedom of Action," *The Philosophical Review*, vol. 87 (1978), pp. 22-40; Peter Van Inwagen, "Ability and Responsibility," *The Philosophical Review*, vol. 87 (1978), pp. 201-24, and *An Essay on Free Will* (Oxford, 1983); David Wiggins, "Towards a Reasonable Libertarianism" in T. Honnrich (ed.) *Essays on Freedom of Action* (London, 1973); Susan Wolf, "Asymmetrical Freedom," *Journal of Philosophy*, vol. 77 (1980), pp. 151-66.

10. A dystonic desire, expressed in Harry Frankfurt's useful terminology, is a first-order desire that one has a second-order volition not to act on. (Frankfurt, "Freedom of the Will," *op. cit.*)

11. Some would argue that we can give a narrower explanation of the subject's bondage. For many hold that motivation by an irresistible desire itself precludes free action, regardless of whether the agent rejects the desire. If this is true, then the compatibilist, of course, can make use of the narrower explanation. But opinion is divided on this issue. See the dispute between Frankfurt and Locke over the case of a drug addict, who is incapable of resisting his addiction, but who is also delighted with his condition and thus does not take drugs only because he is addicted. Frankfurt maintains that, since this addict takes drugs because he wants to, he acts freely. Locke disagrees, maintaining, in effect, that willing slavery is slavery all the same. ("Three Concepts I," *op. cit.*, pp. 98-100, "Three Concepts II," *op. cit.*, pp. 17-19, and "Freedom of the Will," *op. cit.*, pp. 19-20.)

Interesting as this dispute is, its resolution is immaterial to our purposes. Suppose, with Frankfurt, that the willing addict acts freely. This is no comfort to the proponent of the mc argument, since it shows that mc cases modeled on these lines will support premiss 1. Suppose, with Locke, that the willing addict does not act freely. This is no comfort to our opponent either, since it implies that the narrower explanation can be given of MC-A.

12. This is perverse, but possible. Some may be inclined to deny that it is conceptually coherent by arguing as follows: the person induces the irresistible desire in himself because he wants to have it; therefore, when he acts from it, he will not be acting on a desire he rejects. But in fact I may today want myself to do (or to have) something tomorrow that I can anticipate that I will reject the idea of my doing (or having) when tomorrow rolls around. If I now induce an irresistible desire to do it tomorrow, then (assuming that I will be inclined to put up whatever resistance I can), I guarantee that on the next day I will wage a losing battle against the desire.

13. Not all of the circumstances in which one can say, ingenuously, that he is not doing as he really wants are ones in which the agent is in bondage. Frankfurt's "Three Concepts II," *op. cit.*, pp. 113 f., contains some helpful discussion of this issue.

14. Although the analysis of "irresistible desire" is problematic, these words connote immunity to control, inhibition, and so on, by other aspects of the agent's motivational system, including his formed character and value judgments. Once in the sway of such a desire, the agent is unable to stop it from moving him to action even if he has what he regards as the best reasons for doing so and even if he tries his best. On this conception, a desire does not qualify as irresistible merely because it happens to be strongest or because it prevails in volition. Consequently, this notion does not collapse the distinction between free and unfree action.

15. At this point I have found Dennett's discussion of induced beliefs very suggestive (*Brainstorms*, *op. cit.*, pp. 250-53), though I do not know whether the line I have taken here is one of which he would approve.

16. Locke, for example, says the mind has "a power to suspend the execution and satisfaction of any of its desires; and so all, one after the other; it is at liberty to consider the objects of them on all sides, and weigh them with others. In this lies the liberty that man has; and from the not using of it right comes all that variety of mistakes, errors, and faults which run into the conduct of our lives and our endeavour after happiness; whilst we engage too soon, before due examination." Later he adds: "the satisfaction

of any particular desire *can be* suspended from determining the will to any subservient action, til we have maturely examined whether the particular apparent good which we then desire makes a part of our real happiness, or be consistent with it. The result of our judgment upon that examination is what ultimately determines the man; who could not be *free* if his will were determined by anything but his own desire, guided by his own judgment." (*op. cit.*, p. 345 and p. 367, italics in original.)

Suitably modified, the rationality requirement for free action will also cover other ways of rendering MC-B coherent, such as if the controller achieved his ends by inducing a state of insanity. It is unnecessary, I think, to draw out the details of this case or its compatibilist remedy.

17. For some further comments on MC-B", see note 26. - (25)

18. Cf. *Elbow Room*, *op. cit.*, pp. 64-5.

19. Notice that in defending compatibilism I have made use of a central tenet of this doctrine, namely, that reasons are causes. I am aware, of course, that libertarians frequently reject this position. But this is no objection, since the question now at issue is whether compatibilism can handle the mc problem. Naturally, in dealing with the problem the compatibilist will draw on his own views.

20. The requirement of mediation by character can be used to explain the absence of freedom in several kinds of mc cases, including MC-B'. The compatibilist can also use it to cover cases of steady and all-pervasive mc. Imagine case MC-B'", in which the controller manipulates his victim on a continuous basis so that each of the subject's mental states, including his meta-desires, are the product of a specific act of mind control. Here the victim is perhaps as close to being a puppet as we can imagine. In MC-B'" the controller has not merely by-passed the man's character; he has deprived him of any real character of his own. Thus, Frankfurt's remarks about this example are extremely apt: "[the subject's history] is utterly episodic and without inherent connectedness. Whatever identifiable themes it may reveal are not internally rooted; they cannot be understood as constituting or belonging to the subject's own nature . . . [T]he subject has no character or dispositions of his own, and there is no reason to expect from him—except derivatively, insofar as there is reason to expect it from the [controller]—even the minimum of continuity and intelligibility essential to being a person." ("Three Concepts II," *op. cit.*, p. 130)

21. It should be stressed that mediation by one's character is not a sufficient condition of freedom. As I go on to indicate, there are unfree actions that are mediated by one's character.

22. One can make a identical point about efforts to dislodge character traits in adults. Is it inconceivable that someone else should increase my ability to act freely by giving me a drug, administering behavioral therapy, and so on, thereby ridding me of a compulsion and establishing a more benign trait? Naturally, if this were done against my will, then imposing the therapy would violate my autonomy and perhaps would not be justified. But that's a different issue.

23. It is close to the spirit, but not to the letter of Plato. I do not assert, for example, that the free person would necessarily be moral, nor do I commit myself to some of the other metaphysical or political trappings associated with Platonism.

24. Anthony Burgess, *A Clockwise Orange* (New York, 1962) and George Orwell, *1984* (New York, 1949). Alex has been subjected to a process of intense behavioral therapy to rid him of his criminal tendencies. He is repeatedly shown violent films under the influence of a nausea-producing drug and is thus conditioned to experience an overwhelming and incapacitating sense of discomfort at the thought of his doing violence. For an instructive philosophical discussion of *A Clockwork Orange* see P. S. Greenspan's "Behavior Control and Freedom of Action," *The Philosophical Review*, vol. 87 (1978), pp. 225-40.

The fate of Winston Smith is perhaps too well-known to require review. For our purposes, however, it may be important to recall that, although he ends up ardently loving Big Brother, this reaction is wrenched out of him by a process of psychological torture that makes him, in Orwell's words, "helpless, [and] mindless" (p. 235). He lives on in a drunken stupor, psychologically devastated, and robbed of anything vaguely resembling his normal powers of thought and reflection.

25. It may be helpful to contrast this case with MC-B", in which I am not inclined to regard the man as acting freely, despite the fact that his desires are syntonetic, that he has the ability to evaluate his impulses, and so on. If, in MC-C, we have a new person who is free, why not say the same of MC-B"? My intuitions in MC-B" are guided by two features of the case: (a) the man's old character is not obliterated but only repressed and (b) the repression is brief and transitory. It strikes me as more natural to describe this as a case of one and the same individual who does not act freely than of a new person who does. If anyone should have the opposite intuition, that is so much the worse for Argument MC. In any case, I think it is important to show that the compatibilist can accommodate the plausible claim that MC-B" is not a case of free action.

26. I am indebted for helpful comments on earlier drafts of this paper to Laurence Bonjour, Jean Beer Blumenfeld, Harry G. Frankfurt, T. Walter Herbert Jr., Robert Kane, Bernard D. Katz, and Stephen P. Stich.