

IS HARD DETERMINISM A FORM OF COMPATIBILISM?

JEREMY RANDEL KOONS

Libertarianism is ailing. Once the preferred account of free will, it has now fallen into considerable disfavor. Most philosophers are suspicious of libertarian accounts of agent causation and unmoved movers, and are convinced that the opposite of determinism is not freedom, but indeterminism. I think that libertarianism's current low state is entirely justified. I will not argue for this conclusion, as it has been adequately argued elsewhere. This conclusion leaves only hard determinism and compatibilism on the field. Although hard determinism is scarcely more popular than libertarianism, many philosophers seem to reject it, not because of its philosophical implausibility, but because they fear the consequences of its being true. Metaphysicians' inability to come up with a satisfactory compatibilist account is thus a source of worry, and might lead one to fear that hard determinism is true.

On their face, compatibilism and hard determinism are very different. Of course, compatibilists and hard determinists agree that the thesis of determinism is true:

Determinist thesis (D): Every event (including human actions) has a cause, and the chain of causes leading to any given action by an agent extends back in time to some point before the agent was born.¹

Eventually, this chain of causes becomes a question for cosmologists, not for philosophers of free will.

But this is where the similarity between compatibilism and hard determinism is alleged to end. The two theories disagree about the truth of a crucial thesis, the compatibilist thesis:

¹ Of course, it may be that certain quantum events are uncaused, but as these events do not play a significant role in human psychological causation, they may be ignored for the purposes of the free will debate.

Compatibilist thesis (C): Determinism is compatible with free will and moral responsibility.

Compatibilists, who endorse both C and the thesis of determinism, also endorse the freedom thesis:

Freedom thesis (F): Some human actions are free.

Hard determinists, on the other hand, endorse D, but reject C, and consequently must also reject F. Hard determinists realize that humans have a flourishing practice of labeling acts as free and unfree, and granting and withholding ascriptions of moral responsibility, but they think this practice is unjustified.

The goal of this paper is to show that hard determinism actually turns out to be a form of compatibilism. I will accomplish this in three steps. First, I will argue that the actions a compatibilist will punish or reward tend to fit a certain criterion: they tend to have a particular feature. This criterion will not amount to a definition of “free action,” but is instead a defeasible criterion licensing the following, defeasible inference: if an action falls under this criterion, then, *prima facie*, the action is free. Second, I will show that hard determinists will end up punishing or rewarding action (and inaction), and the actions they punish and reward will also tend to fall under this criterion. Thus, we will see that the hard determinist endorses the following thesis:

Sanctions thesis (S): Some human actions are sanctionable;

and that “sanctionable actions” and “free actions” have the same extension, that is, they pick out the same set of actions.

Finally, I will respond to some objections. At this point, it will emerge not only that the same actions will be regarded as sanctionable by the hard determinist and free by the compatibilist (so that the two theories are equivalent at the practical level), but also that we ought to translate “sanctionable” as “free.” Thus, the hard determinists end up endorsing the compatibilist thesis (some human actions are sanctionable/free), and theses F and S turn out to mean the same thing. Thus, hard determinism is quite literally a form of compatibilism. Therefore, assuming that libertarianism is false, compatibilism is the only remaining account of free will, and must be the one we endorse.

FREE ACTION

Let me begin by expressing some doubt regarding metaphysicians’ attempts to give a one-size-fits-all definition of such concepts as *free will* and *personal identity*. First, it is not clear that our concepts have sufficiently sharp boundaries to

admit of such explicit definition. Consider Wittgenstein's point that there is nothing had in common by all instances of our application of the term "game."² The terms we use generally cannot be defined in a way that captures all correct applications of the term. Nor is it clear that we need to give such explicit definitions to our terms. It might ease application, but it's not clear that the benefits of this outweigh the disadvantages of having to use the resulting Procrustean concepts, with their awkward, contrary-to-practice contours and counterintuitive results. Thus, I do not think we need to worry if we cannot give a set of explicit rules for the application of terms such as "free action" (i.e., "An action A is free if and only if . . ."). A related lesson is gaining prominent attention in ethics, where particularists of various stripes argue against the existence of exceptionless moral principles.³

We should thus expect that we will not be able to give explicit and exhaustive definitions for most of our concepts. Instead, we may only be able to give criteria, or rules for the application of a concept, which will license defeasible inferences of the form "If action A has feature F, then, prima facie, action A is free."⁴ In other words, there might be times when correct application of a term violates the explicit rule we have provided for the application of this term (much as the duck-billed platypus violates the rule that the term "mammal" is applied only to animals that give birth to live young). With this in mind, I now turn my attention to providing a criterion for the term "free action."

Consider the case of a child drowning in a shallow pool. Person A walks by and, deciding he doesn't want to ruin his slacks, declines to wade into the pool and rescue the child. We condemn his inaction as evil. Consider, though, person B, who is chained to a nearby tree and helpless to rescue the child. No one would blame person B for his inaction. What is the difference between these two scenarios that causes us to condemn person A and not person B? It cannot be that person A could have done otherwise, whereas person B could not have. Given the set of physical laws P and the circumstances C immediately prior to A's inaction, there is only one behavior that could have been exhibited by person A: namely, inaction. This is actually a matter for debate, but for rhetorical reasons, I wish to concede this point to the hard determinist, for I believe that my argument works even after I make this concession regarding determinism and necessitation. In other words, there are no possible worlds in which, given the

² Ludwig Wittgenstein, *Philosophical Investigations*, trans. G. E. M. Anscombe (New York: Macmillan Publishing Co., Inc., 1958), 31–32.

³ For the latest on this debate, see Brad Hooker and Margaret Olivia Little, eds., *Moral Particularism* (Oxford: Clarendon Press, 2000).

⁴ Important work on defeasible generalizations (that is, rules that have genuine authority, and which are not mere statistical generalizations, but which nevertheless allow of exceptions) is being done by Mark Lance and Margaret Little.

same circumstances and physical laws, person A would have done otherwise than he actually did. So A could not have done otherwise, any more than B could have done otherwise. What, then, is the difference? The difference seems to be that A's behavior is the sort of behavior that could be altered through education and application of sanctions. In other words, if we could somehow get A to grasp the suffering of the child's family, or if we could offer him praise for rescuing the child or punishment for failing to rescue the child, then his psychology might be altered in a way such as to cause him to rescue any future drowning children he might encounter. However, no amount of education, praise, or punishment could alter the behavior of the man chained to the tree. In other words, "failure to rescue a child by reason of being chained to a tree" is not the sort of behavior that can be altered by education or sanctions; "failure to rescue a child by reason of not wanting to ruin your slacks" is behavior that can be reshaped through appropriate means.

Or consider the case of a weak-willed man who often drinks a third glass of wine at parties when his limit is two. His behavior is the result of an internal struggle between his moral desires and his desire to drink more alcohol. We could change his behavior by any number of means. We could show him statistics about how alcohol impairs one's ability to drive, or show him grisly photographs of car accidents. Or we could fine him for driving under the influence. Or we could, as a group, ostracize him until he stops drinking and driving. Any of these courses of action has, as a possible outcome, alteration of the man's future behavior. His is the *type* of behavior that can be reshaped in this way. However, punishing the man chained to the tree for not rescuing the child will not affect the chances of his rescuing a child when next chained to a tree in full sight of a drowning child. His is not the *type* of behavior that is amenable to such alteration. Thus, we hold the incontinent man responsible for his actions, but not the chained man.

Of course, the fact that a certain type of action or inaction might be altered by sanctions and rewards does not mean that that action should be blamed; it just means that the person is responsible for this action, good or evil. Consider person C, a wheelchair-bound quadriplegic who observes the child drowning and does not enter the water. Perhaps some extreme form of social sanctioning could alter the behavior of person C, causing her to enter the water. But any such attempt to rescue the child would lead only to the death of C, as well, and would not benefit the child in any way. So, if we do label C's action as free, then we ought to praise her inaction, for entering the water would surely have been foolhardy. Similarly, a healthy person who entered the water and saved the child is to be praised: although her action is the sort that is susceptible to "engineering," it is also the sort of action that we want to encourage. If this same person were being remote-controlled by a mad scientist, though, we would not praise her, because praising her would have no effect on future cases of being under control by a mad

scientist. We might praise the scientist, though, for directing this person to rescue the child (although we might instead punish the scientist for placing people under remote control).

It thus appears that free actions share a common feature. A behavior is free if and only if it is the type of behavior that is amenable to alteration by reward, punishment, retraining, etc. Application of moral praise and blame is part of this “moral engineering”: it reinforces desirable behavior and provides a disincentive for undesirable behavior. This criterion seems neatly to capture the contours of the concept of freedom (and hence responsibility). Considerations distinct from this criterion might justify us in applying the term “free” in a case where the criterion is violated, or in withholding application of the term in a case where the criterion is satisfied. But this is the way it goes with defeasible criteria and entailments.⁵ Let us now look at how a community of hard determinists might apply praise and blame in a way that follows the contours of this criterion for freedom we have provided.

HARD DETERMINISM AND FREEDOM

In order to defend my thesis, I will tell a fable about a community with a particularly active and committed intellectual life. Let us imagine a community of philosophically minded citizens who are not afraid to bite the bullet in their quest to live life according to their philosophical convictions. We can further imagine that this community, embracing both determinism and incompatibilism, decides to jettison the term “free will” and quit applying moral praise and blame. Suppose further that the worst fears of the foes of hard determinism are realized: these citizens, freed of moral responsibility and without threat of punishment to keep them in line, cease to observe the constraints of conventional morality and begin to murder, steal, rape, and pillage at will. The town elders, seeing that this is a disaster which is reducing the quality of life for everyone, decide that even if there is no freedom, there needs to be a certain amount of social engineering to keep society well-ordered and smooth-running. They thus decide to punish actions which hinder the well-ordering of society (murder, theft, etc.) and reward actions which promote order in society (honesty, benevolence, etc.). This works reasonably well; the incidence of lying, murder, theft, and so on decreases dramatically. There are, however, a few cases in which this strategy doesn’t work. Punishing people who, because they are chained to a tree, fail to rescue drowning children has led to no decrease in the incidence of inaction in similar circumstances. The elders meet again and decide that such behavior simply isn’t

⁵ It will emerge below that my account relies on the notion of a practice, consisting of a set of rules. These rules will, of course, also have to be defeasible.

the sort of behavior that is amenable to alteration by punishment. Further, since punishing these people is costly and yields no benefits to society (since this punishment doesn't alter future behavior of people chained to trees), it is decided that such behavior should no longer be punished. Thus, the elders decree the following: a behavior is sanctionable if and only if it is the type of behavior that is amenable to alteration by reward, punishment, retraining, etc. The elders are pleased with this decree, because it has some pleasing results. Whereas before, they were rewarding people who, under remote-control from mad scientists, rescued drowning children, they are no longer obliged to reward such people. One of the elders points out that such rewards were an unnecessary expense, since they had no effect on whether the person in question performed future rescues while under remote control. Rewarding the mad scientists, however, does seem to have a positive effect on the instance of remote-controlled life-saving (although the scientists are later punished when it is discovered that in general, the people of the community don't want to be remote-controlled—this is a result that later turns out to be important⁶).

Thus, the hard determinists see the value of socializing people to certain sorts of behavior. They further realize that the application of praise and blame is a much more effective method of social control than a Hobbesian leviathan. So they encourage each other to resume labeling acts as right and wrong, and punishing those who commit wrong acts and rewarding those who commit right acts. Of course, an action must be sanctionable before sanctions are warranted, so only those actions which fit the given criterion for sanctionability are subjected to praise and blame. But this will have the result that our hard determinists will become compatibilists again. Consider the similarity between the hard determinist thesis and the compatibilist thesis. The compatibilist thesis is the following:

A behavior is free if and only if it is the type of behavior that is amenable to alteration by reward, punishment, retraining, etc.

The hard determinist thesis is this:

A behavior is sanctionable if and only if it is the type of behavior that is amenable to alteration by reward, punishment, retraining, etc.

Eventually, then, the behavior of our hard determinists becomes quite similar to that of our compatibilists. If we were to find ourselves among these hard determinists and had to write a translation manual, it is not beyond the realm of

⁶ See, for example, the section "Draconian Sanctions?"

possibility that we would translate the term “sanctionable” as “free” (although I will return to this point anon⁷). Hard determinism is beginning to look a lot less like a brave new world, and more like business as usual.

The Thesis Generalized

It seems likely that hard determinism and compatibilism will coincide regardless of what theory of compatibilism you endorse. That is, the account of free will that is most rational for the compatibilist to endorse will likely also be the one that is best for the hard determinist to adopt. What makes this likely is that what is at stake in the free will debate is responsibility, typically moral responsibility. Acts that are free are acts for which we are responsible; and we are not responsible for unfree acts.⁸ If you are responsible for an action, then you are the one praised or blamed for this action. Praise and blame, in turn, serve to reinforce behavior we value and discourage behavior we do not value. But this is precisely why the hard determinist is distributing praise and blame: to encourage certain behaviors and discourage others. So it appears that the hard determinist and the compatibilist will end up sanctioning the same behavior. Consider this: the compatibilist will want to punish those free acts which are wrong, or which exhibit traits that we devalue, and reward those which are right, or which exhibit traits we value. The hard determinist will want to select those acts for which we are responsible, and either reward or punish these acts, depending on the hard determinist’s system of values. Since free acts and acts for which we are responsible are the same, then the compatibilist and the hard determinist will end up sanctioning the same behaviors. (This assumes, of course, that the hard determinist community has the same interests and values as the compatibilist community. We will return to this point later.)

Indeed, the compatibilist intuition is that an action is free if it was caused in the right way by an agent’s desires; the trick is to flesh out the phrase “in the right way.” But desires are the sorts of things that are subject to training; so, given this compatibilist intuition, free actions are going to be the sorts of actions that can be encouraged or discouraged through the application of sanctions or rewards. But the hard determinist is sanctioning exactly this set of actions. Thus, it seems likely that a successful compatibilist account will have contours similar to the contours endorsed by our community of hard determinists.

⁷ See the section “Are We Free or Are We ‘Free’?”

⁸ Harry Frankfurt thinks that freedom and responsibility come apart, but I think that he is driven to this position by some of the counterintuitive results of his overly formalized definition of free will. See Harry Frankfurt, “Three Concepts of Free Action: Part II,” *Proceedings of the Aristotelian Society*, Supp. 49 (1975), 113–25.

Pragmatism, Freedom, and Hard Determinism

The case for thinking that the hard determinist and the compatibilist will end up sanctioning the same behaviors is particularly strong if the compatibilist is a pragmatist about free-will talk.⁹ I think the compatibilist should be a pragmatist about free-will talk; elsewhere¹⁰ I argue in greater detail that varieties of normative discourse (moral evaluation, epistemic evaluation, etc.) must be justified pragmatically, and ascriptions of free will seem to be a species of normative discourse. But if the compatibilist is a pragmatist about free-will talk, then the compatibilist and the hard determinist will both be approaching the free-will debate in exactly the same manner: both will recognize that there are certain legitimate ends (such as causing behavior to conform to certain moral norms, which are themselves justified pragmatically by appeal to peaceful and fruitful coexistence, etc.) that can be advanced successfully by instituting a practice of praise and blame, punishment and reward. There is no reason in particular to think that the practice of praise and blame that serves these legitimate ends best in the compatibilist community would be different from the one that serves these ends best in the hard determinist community; after all, the same ends are being served by a pragmatically justified practice of free-will ascriptions.

Consider the matter this way: for a pragmatist, varieties of normative discourse are structured so as to serve our interests. Moral norms serve our interest in peaceful and fruitful coexistence; epistemic norms serve our interest in truth, explanation, and instrumental control over the environment; and so forth. There is no reason why the hard determinist community should have different interests from ours: they, too, want to live together peacefully and fruitfully; they, too, have an interest in understanding, explaining, and controlling the world around them; they, too, have an interest (arising out of sympathy) for seeing their friends and neighbors flourish. They will even have an interest in liberty: although they might reject the metaphysical notion of freedom, they will not want to be deprived of political and social liberty, which is the right to pursue their interests and prefer-

⁹ To be a pragmatist about a type of discourse is not merely to suppose that we pursue the discourse for pragmatic reasons; it is possible that we pursue science for pragmatic reasons, but this shouldn't force us to be pragmatists about science. Rather, pragmatism has to do with the relation between our interests and the truth of the claims that constitute the discourse: it is true that electrons have a negative charge because electrons have a negative charge; but it is true that murder is wrong because forbidding murder serves a number of our interests. Viewed this way, free-will talk is a sort of second-level practice: it piggybacks on our moral and other types of practices, and is instrumental in their execution.

¹⁰ See Jeremy Randel Koons, "Do Normative Facts Need to Explain?" *Pacific Philosophical Quarterly* 81 (September 2000): 246–72.

ences more or less free from coercion by others.¹¹ If they have the same interests, then we have no reason to suppose that the varieties of *pragmatically justified* discourse (normative, epistemic, freedom/sanctionability, etc.) will differ between the two communities. Both communities will participate in the same pragmatically justified (but not necessarily *utilitarian*¹²) practices.

So it seems that a true theory of which acts we are responsible for (i.e., which acts are free) will serve equally well as a compatibilist account of freedom and responsibility and a hard determinist account of freedom and responsibility. Thus, the result I am defending does not depend particularly on the account of free will I sketched at the beginning of the paper.

OBJECTIONS TO THE THESIS

I now must turn to several objections that attempt to pry apart compatibilism and hard determinism. I will argue that these objections do not disprove my general thesis, and that professed hard determinists are still squarely in the compatibilist camp. Responding to these objections will give me further opportunity to demonstrate that the hard determinist and compatibilist communities will behave in the same way. It will also emerge that “sanctionable” and “free” mean the same thing; this will allow us to conclude that hard determinism genuinely is a form of compatibilism.

Retributive Punishment

One alleged difference between the hard determinist and the compatibilist is that while the compatibilist might endorse purely retributive (as opposed to corrective) punishment, the hard determinist will not employ this form of punishment. After all, the hard determinist punishes in order to alter future behavior or to present a deterrent to would-be miscreants, but purely retributive punishment seems an unnecessary expense for the community of hard determinists.

¹¹ Even if the hard determinists value liberty, one might still object that they do not value freedom (in the metaphysical sense), and so there is a point at which the interests of hard determinists and compatibilists diverge. In the section “Is Freedom Intrinsically Valuable?” I argue that both hard determinist and compatibilist communities might value freedom/sanctionability only instrumentally; and so their interests will not diverge at this point, either.

¹² It is important to note that *pragmatist* is not the same as *utilitarian*. A pragmatically structured practice is, of course, structured so as to serve our interests, but this becomes a utilitarian account only on the assumption that there is some quantifiable stuff, “goodness,” which such a practice ought to maximize. I see no reason to think that this assumption is true; and so the account of freedom/sanctionability offered here is a pragmatist account, but not a utilitarian one. (I am grateful to Mark Lance for helpful correspondence on this topic.)

The hard determinist might endorse non-corrective punishment to stand as an example to others, and thereby act as a deterrent, but many compatibilists will say that evil deserves punishment, even if no one ever knows about the punishment (which would, of course, preclude the possibility of this punishment serving as a deterring example). Evil actions, they will say, simply deserve to be paid back.

There are two possible responses to this attempt to distinguish hard determinism and compatibilism. First, one might claim that failure to endorse the practice of purely retributive punishment hardly puts one outside of the compatibilist community. Many claim that retributive punishment panders to the lowest of human impulses, the desire for revenge, and that it has no place in civilized society. Punishment should be purely corrective, these people will argue; non-corrective punishment is pointless infliction of suffering. Indeed, we often seem willing to give a second chance to those who have turned their lives around, or who seem genuinely remorseful for what they have done. It seems, then, that our commitment to retributive punishment is not as strong as one might think.

The second (and, I think, more plausible) response is this: the practice of retributive punishment might be necessary in a hard determinist community. Suppose there is no purely retributive punishment. Then people know they can get away with single, unrepeatable infractions, such as killing their only relative (who happens to be wealthy) to get his inheritance. No corrective punishment is required, since the murderer has no more relatives, and recidivism is already ruled out. But a society that practices purely retributive punishment is also a society in which people know they cannot get away with even these unrepeatable infractions. Thus, retributive punishment serves a role in the ordering of society, and hence can be practiced by the hard determinist.¹³

Indeed, this is a point worth dwelling on. For a practice of punishment to serve the goals of the hard determinist community, the members of the community must associate guilt with punishment: unless you know that punishment follows guilt, the punishment serves as no disincentive to incur guilt. But a practice of purely retributive punishment makes the connection between guilt and punishment a strong one, and would hence well-serve the hard-determinists' goal of deterring wrong behavior.¹⁴ It is clearly possible, then, for compatibilists and hard determinists to be on the same wavelength regarding retributive punishment. So non-corrective punishment cannot be the wedge that pries the two theories apart. More remains to be said about the practice of retributive punishment. A natural place to say it is in response to the next objection.

¹³ I will address the question of whether such punishments in the hard determinist community really ought to be considered *retributive* punishments in the section, "Are We Free or Are We 'Free'?"

¹⁴ For a discussion of practices and the pragmatic justification for adopting practices, see John Rawls, "Two Concepts of Rules," *The Philosophical Review* 64 (1955), 3–32.

Punishing the Innocent?

A persistent worry about justifying retributive punishment pragmatically is that, according to the hard determinist, it seems to follow that if it would serve some social purpose to “punish” a person who had committed no crime, then this would be permissible. The compatibilist, of course, might well say that one cannot punish an innocent person if he or she has not committed a crime, even if so punishing the person might serve some societal purpose.

However, it is not clear that the hard determinist must allow punishment of the innocent, even if doing so would serve a pragmatic purpose. To say that one justifies a practice (such as the practice of retributive punishment) pragmatically is not to say that one can appeal to these pragmatic considerations to justify individual actions.¹⁵ The reasons that justify the practice are not allowed to “trickle down” to the level of individual acts; the practice ceases to serve its purpose if that is allowed. This fact was seen clearly by Rawls in his 1955 article, “Two Concepts of Rules.” In this article, Rawls tries to argue that rule utilitarianism is superior to act utilitarianism. With the latter, utilitarian considerations justify individual actions. However, since allowing utilitarian considerations to justify individual actions leads to free rider problems and the like, act utilitarianism fails to maximize utility. With rule utilitarianism, however, utilitarian considerations justify a practice (consisting of a set of moral rules), but individual actions are justified by appeal to these rules, and not by appeal to utilitarian considerations. As Rawls writes:

There are obvious utilitarian advantages in having a practice which denies to the promisor, as a defense, any general appeal to the utilitarian principle in accordance with which the practice itself may be justified. There is nothing contradictory, or surprising, in this: utilitarian (or aesthetic) reasons might properly be given in arguing that the game of chess, or baseball, is satisfactory just as it is, or in arguing that it should be changed in various respects, but a player in a game cannot properly appeal to such considerations as reasons for his making one move rather than another. It is a mistake to think that if the practice is justified on utilitarian grounds then the promisor must have complete liberty to use utilitarian arguments to decide whether or not to keep his promise. The practice forbids this general defense; and it is a purpose of the practice to do this. Therefore what the above arguments [against utilitarianism] presuppose—the idea that if the utilitarian view is accepted then the promisor is bound if, and only if, the application of the utilitarian principle to his own case shows that keeping it is best on the whole—is false. The promisor is bound because he promised: weighing the case on its merits is not open to him.¹⁶

So, even if we acknowledge that pragmatic considerations justify a practice, we are not forced to admit that these same considerations can be appealed to in

¹⁵ In what follows I draw from my article, “Do Normative Facts Need to Explain?” where I argue in more detail that pragmatic considerations are not allowed to justify individual actions within a pragmatically justified practice.

¹⁶ Rawls, “Two Concepts of Rules,” 16.

justifying individual acts within the practice, any more than we have to concede that a chess or basketball player can appeal to pragmatic or utilitarian considerations to justify a particular play. Thus, the objection that pragmatically justifying the practice of punishment will lead to the punishment of innocents fails to recognize the difference between the pragmatic justification of a practice and the pragmatic justification of an individual action. The former is permitted; the latter is forbidden, for allowing the reasons that justified the practice to seep *into* the practice and justify individual actions causes the practice to collapse, rendering it incapable of serving its purpose.¹⁷

Is Freedom Intrinsically Valuable?

Many of us think that freedom is intrinsically valuable: that we should be allowed to be free, even if this freedom leads to bad results. Indeed, a standard response to the problem of evil is to say that a world of people with free will, and who therefore perform some evil acts, is better than a world of people with no free will, but who are constrained to perform only good acts. It's not clear, though, that endorsing compatibilism requires that we place intrinsic value on freedom. First, nothing in the definition of compatibilism says that autonomy is intrinsically valuable. Second, it is not clear that we do place intrinsic value on freedom. To take a related example, one of the classic defenses of political liberty, John Stuart Mill's *On Liberty*,¹⁸ argues that we ought to allow liberty for the benefits liberty brings (societies that progress more quickly than do conformist societies, the contribution of new truths to our belief system, the strengthening of old truths by their clash with false views, etc.). Nowhere does Mill defend the claim that liberty is intrinsically valuable. It seems likely that the compatibilist can pursue a similar line regarding metaphysical freedom. For these two reasons, I don't think that philosophers need to think that freedom is intrinsically valuable for them to be allowed to count as compatibilists.

Thus, if our hard determinists place only instrumental value on freedom, this doesn't preclude them from being compatibilists. And I think that the hard determinist can endorse the type of defense of freedom that Mill gives for liberty. A free society is a better and more pleasant place to live; we have an interest in living free from coercive authority. Thus, while the way of thinking which leads

¹⁷ One might nevertheless object that if the retributive practice is given a pragmatic justification, then the punishment dictated by this practice is not really retributive punishment, but is instead deterrent punishment; and so one might maintain that the hard determinist will not, in fact, punish retributively. I will return to this objection in the section, "Are We Free or Are We 'Free'?"

¹⁸ John Stuart Mill, *On Liberty*, ed. Elizabeth Rapaport (Indianapolis, IN: Hackett Publishing Company, Inc., 1978).

philosophers to embrace hard determinism might lead them to reject placing intrinsic value on freedom, this rejection is not in itself incompatible with compatibilism.

In any case, I suspect that members of the hard determinist community would place some intrinsic value on freedom. Think about what it would mean for there to be no sanctionable behavior in the community, that is, no behavior amenable to alteration by punishment or reward. That would require that some authority was exercising such massive control over people's psychology that these people's behavior would cease to be alterable by the application of sanctions and rewards. I doubt that the members of the hard determinist community would find this idea very appealing; they would not want (any more than we would) to cede such control or authority to anyone. They would like and value sanctionability. Hence, this worry about the value of freedom/sanctionability seems not to be justified.

Draconian Sanctions?

Another objection that might be raised against my thesis is the claim that the sanctioning behavior of the hard determinist community would actually be quite different from that of the compatibilist community. Hard determinists would be less averse to forms of control (brainwashing, etc.) which compatibilists find abhorrent. Thus, hard determinism isn't really just compatibilism by a different name.

I think that this objection also has a response. Remember why the hard determinists reinstated freedom talk and responsibility talk. Order was breaking down, and the members of the community didn't like this. They reinstated social controls because they liked order, for order brings security and allows us to pursue the things we like and value. In short, the hard determinist community will discourage behavior which they dislike or devalue (or which produces things they dislike or devalue), and encourage behavior which they like or value (or which produces things they like or value).

But is this different from what we do now? We encourage moral behavior because we value it; we encourage hard work and industry because these are traits we value. Our distribution of praise and blame, punishment and reward, is based on our values and preferences. This is no different from the pattern used by the hard determinists: the distribution of sanctions reflects their values. So, assuming the two communities value the same things and have the same interests (and why wouldn't they?), they will encourage the same sorts of behaviors. These behaviors will include sanctioning behaviors. In other words, *we sanction the methods others use to sanction* if we think that these people are going about it the wrong way. We may disapprove of pornography, but think that firebombing a pornographer's house is an inappropriate punishment. We might disapprove

of theft, but think that allowing *Clockwork Orange*-style behavioral modification to be practiced on such individuals gives too much power to too few individuals or makes us fear the government; or perhaps our sense of sympathy and fellow-feeling condemns it. (The reader should be reminded of our discussion of the value of sanctionability at the end of the previous section.) And this is why the hard determinist community won't necessarily be a 1984-style police state, with mind control and brainwashing. The community sanctions behavior to produce behavior it is willing to tolerate, and if the community is unwilling to tolerate draconian punishments (and clearly, we don't tolerate these sorts of punishments), then this sort of sanctioning behavior will not be tolerated. The sanctioning behavior of the hard determinist community will be determined by its interests. And since (as I argued earlier) the hard determinists' interests will be substantially the same as those possessed by a compatibilist community (up to and including interests in liberty and aversion to coercive authority), the members of the hard determinist community will not permit such objectionable punishments.

Freedom and the Mentally Ill

There might be a difference between the behavior of our hard determinists and that of our compatibilists. Consider the criminal behavior of one who is mentally ill, but whose illness might be treated through medication or therapy. This criminal's behavior fits the criterion for "sanctionable behavior," but compatibilists would hesitate to call the action free. Indeed, juries often are given the option of ruling a defendant not guilty by reason of insanity. Such people are thought not to be responsible for their actions. In short, they are not free, although the hard determinist might call their actions sanctionable. What this means is that the hard determinist will be sanctioning actions we regard as unfree; and this might be thought a salient difference between the two positions.

An important thing to note is that, even though we do not call such people free, we sanction them all the same. We commit them (typically against their will) to mental hospitals, and force them to undergo therapy with the goal of altering their behavior. The difference seems to be that, with violators found not guilty by reason of insanity, our treatment is aimed at rehabilitation, not retribution, whereas our treatment of violators we consider sane is aimed at both punishment and improvement.

Notice, though, that if the compatibilists did away with retributive punishment and aimed all treatment of violators at rehabilitation, then their behavior would be brought into line with that of the hard determinists who make no distinction, in terms of labeling actions as sanctionable, between the sane and the criminally insane. Both the compatibilists and the hard determinists would then focus solely

on rehabilitation. If the compatibilists did away with retributive punishment, such an action would probably be accompanied (or even caused) by the realization that criminals are poorly socialized, or often suffering from mental or physiological disorders, in much the same way that the criminally insane are. If this were the case, then the rationale behind calling the actions of one group free, while calling those of the other group unfree, would be removed, and the compatibilists might begin regarding the actions of the two groups as the same.

Remember, though, that the hard determinist need not give up on retributive punishment. As we discussed above, the same considerations that might cause the compatibilists to employ retributive punishment could also persuade the hard determinists to do the same.

In any case, the important thing to remember is that the hard determinists' and the compatibilists' treatment of the mentally ill is not significantly different. Both treat the mentally ill with an eye to rehabilitation, or, if rehabilitation proves futile, then both will at least insure that the person in question is prevented from doing any harm. Thus, this case cannot be the source of any important differences between the hard determinist and the compatibilist.

As an addendum, I should note that similar comments apply to the two communities' treatment of children. A three-year-old child is regarded (in a compatibilist community) as in general not morally responsible for her actions. Thus, one might object that since the behavior of this child might be affected by aversive therapy, the hard determinist might well punish the child, whereas the compatibilist would not.¹⁹ Two things are to be noted at this point: first, we *do* punish three-year-old children. We give them "time outs," or send them off to bed early, or revoke certain privileges, etc. Second, there is no reason why the hard determinist community must use punishment as its only method for discouraging bad behavior. The hard determinist has the same concerns as the compatibilist with regard to such a child: a desire not to cause the child to suffer, when causing such suffering is not necessary to morally educate the child; concern for the child's psyche and future development; and so forth. So the hard determinist might well choose education over punishment. Thus, the hard determinist and the compatibilist need not differ in their treatment of children.

Is Punishment Permissible in a Hard Determinist Community?

It has occasionally been suggested that punishment is unjust, if incompatibilism and determinism are both true. Thus, our hard determinists ought not punish people for stealing, killing, etc., because these people aren't really free, and hence are not morally responsible. It's not clear that this objection is coherent, though,

¹⁹ An anonymous referee for *The Philosophical Forum* raised this objection.

for the critic is caught on the horns of a dilemma. On the one horn, if we may with justification say that the hard determinists ought not punish people, then we are admitting that the hard determinists are subject to moral evaluation and moral constraint. But if this is the case, then they are justified in punishing people, for the very reason that they are subject to moral evaluation and moral constraint. So the objection defeats itself the moment it is spoken. On the other horn, if the hard determinists are not subject to moral constraints, then they can hardly be morally constrained not to punish people according to their criterion for “sanctionability.” So either way, the hard determinists may punish people without being subject to moral criticism.

However, as I will argue in the next section, the hard determinists are free and morally responsible in precisely the same sense as are compatibilists. We will see that “sanctionable” means the same thing as “free,” and “moral responsibility” as spoken by a hard determinist means the same thing as “moral responsibility” as spoken by a compatibilist. Therefore, there should be nothing morally objectionable about punishment in a hard determinist community.

Are We Free or Are We “Free”?

There is one final objection to be met. Suppose I have shown that, at the practical level, compatibilism and hard determinism are identical, that the two communities behave identically. One might nevertheless wonder whether it follows that hard determinism is a *form* of compatibilism; that is, one might wonder whether the word “sanctionable” as spoken by a hard determinist means the same thing as “free” spoken by a compatibilist. An example will illustrate this worry.²⁰ Suppose Smith believes that morality consists of following God’s commands (call this “God worship”), whereas Jones holds that morality consists of following the patterns of nature (call this “nature worship”). Suppose, further, that what Smith believes God to have commanded is exactly what Jones takes the patterns of nature to be. Thus, their behavior is the same. But it clearly does not follow that nature worship is a form of God worship. Similarly, one might claim that though the hard determinists behave the same as the compatibilists, hard determinism is still not a form of compatibilism.

Let me begin by noting that on many accounts of meaning, the question “The two communities’ behavior is identical, but do their words mean the same thing?” is deeply confused, and reveals a misunderstanding of what meaning-talk is all about. For these philosophers, then, “sanctionable” and “free” have the same meaning, since the two words are used identically in the respective communities.

²⁰ The following example was suggested by an anonymous referee for *The Philosophical Forum*.

Nevertheless, I think this objection is worthy of closer examination than this. We will examine some key terms, and it will emerge that these terms mean the same thing when spoken by the compatibilist and the hard determinist. So, while “nature worship” and “God worship” might mean different things, we will see that “sanctionable” and “free” have the same meaning. Other key terms (“retributive punishment” and “moral responsibility,” in particular) mean the same thing when spoken by the hard determinist and the compatibilist. Let us consider these terms one by one.

Sanctionable vs. free: The worry here is that “sanctionable” as spoken by a hard determinist doesn’t have the same robustness as does the word “free” spoken by the compatibilist. I think this is not true. I think the truly robust sense of freedom (freedom from all compulsion, external and internal; ability to do otherwise; etc.) was lost along with libertarianism. What remains for the compatibilist is a more modest conception of freedom: a free act just is an action that is a candidate for reward and punishment, etc. Not only does this definition pick out the actions that are free, it is not clear how the phrase “free action” could mean anything more than what I have here used to characterize the actions it picks out. But all of this is precisely what characterizes the hard determinist’s notion of freedom, as well. For the hard determinist, a free action is just one that is a candidate for reward and punishment, etc. I can see no compelling rationale to refuse to translate “sanctionable” as spoken by a hard determinist as “free” in the mouths of the compatibilists.

However, hard determinists and compatibilists might have auxiliary commitments that will cause “sanctionable” and “free” to diverge in meaning. Let us examine some other important terms connected to the notion of freedom/sanctionability and see if these words have different meanings for the two communities.

Moral responsibility: Different sets of pragmatically justified norms can be classified according to which interests they serve. So, our epistemic norms are those that serve our interest in truth and explanation. Semantic norms are those that serve our interest in communication and forging linguistic bonds between disparate communities. By contrast, our moral norms are those that serve our interest in peaceful and fruitful coexistence.

So far, we have encountered no reason to suppose that the interests of the hard determinist and compatibilist communities will diverge. And so each community will have a set of norms characterizable as moral norms, norms aimed at promoting peaceful and fruitful coexistence. Assuming similarity of interests between the two communities, these norms will themselves be similar between the two communities. So the only question remains, “What does it mean to be *responsible* vis-à-vis this set of norms?”

Well, according to a standard definition, “moral responsibility” refers to “the fulfillment of the criteria for deserving blame or praise (punishment or reward)

for a morally significant act or omission.”²¹ But of course, “sanctionable” and “free” are both just shorthand for “deserving blame or praise.” And so “morally responsible” will mean, *for both communities*, “free/sanctionable with respect to norms of morality,” which norms (as we have seen) do not differ between the two communities.²² Thus, the phrase “moral responsibility” appears to have the same meaning in both communities.

Retributive punishment: I have argued above that the hard determinist community will, on pragmatic grounds, institute a practice of retributive punishment. One might argue, though, that punishments doled out in accordance with this practice are not really retributive punishments, as the purpose of such punishments is (essentially) to deter wrong-doers.²³

This is a serious objection, but ultimately one that can be met. Recall our earlier discussion of practices. Rawls emphasized that the considerations that justify the practice are not allowed to “infiltrate” the practice and justify individual acts within the practice. Individual acts are *essentially* justified by the rules of the practice. As Rawls puts the point,

It is a mistake to think that if the practice [of, e.g., promise-keeping] is justified on utilitarian grounds then the promisor must have complete liberty to use utilitarian arguments to decide whether or not to keep his promise. The practice forbids this general defense; and it is a purpose of the practice to do this.²⁴

Thus, when the hard determinist punishes a person retributively, she *must* punish this person *merely because* the person deserves to be punished—merely because the rules of the practice demand it. No act of retributive punishment is justified by considerations of deterrence or social utility; allowing such considerations to seep into the practice causes the practice to collapse, rendering it incapable of serving its purpose. And so, for the hard determinist, retributive punishments genuinely are retributive.

To conclude this section, we have seen that at the end of the day, the hard determinists and the compatibilists have the same linguistic commitments. Terms such as “moral responsibility” and “retributive punishment” have the same meaning in the two communities, and “sanctionable” has the same meaning as

²¹ Martha Klein, “Responsibility,” in *The Oxford Companion to Philosophy*, ed. Ted Honderich (Oxford: Oxford University Press, 1995), 771–72, quotation on p. 772.

²² Of course, the hard determinists might not use the phrase “moral responsibility,” having abandoned it (along with “free”) in the initial stages of our myth. Obviously, though, what word they use is irrelevant; we are concerned with the meaning of whatever word they end up using to take its place.

²³ An anonymous referee for *The Philosophical Forum* raised this objection.

²⁴ Rawls, “Two Concepts of Rules,” 16.

“free.” Thus, the hard determinists and the compatibilists are not like Smith and Jones; hard determinism *really is* a form of compatibilism.

CONCLUSION

Whereas the compatibilist will answer the question, “Are some actions free?” in the affirmative, the hard determinist will give an affirmative response to the question “Are some actions sanctionable?” Since, as I have argued, the behaviors of the two communities will coincide; and since, as I also argued, we should translate “sanctionable” as “free,” the hard determinist will, it seems, be a variety of compatibilist. Certainly, the typical hard determinist would deny this. But such a blanket denial is insufficient; the hard determinist must be able to point out salient differences between his position and that of the compatibilist. I have argued that such a difference cannot be found, even if we look at such likely candidates as retributive punishment and the manner of sanctions allowed by the hard determinist community. It is possible that the hard determinist may end up being an atypical compatibilist, and endorsing views that few compatibilists would be comfortable holding. But, as we have seen, these particular views don’t suffice to push the hard determinist out of the compatibilist camp. I must conclude, then, that hard determinism is a form of compatibilism.

American University of Beirut, Lebanon