

## Compatibilism, Common Sense, and Prepunishment

Matthew Talbert  
West Virginia University

Published in *Public Affairs Quarterly* 23 (2009): 325-335

Compatibilism is committed to the claim that even if we live in a deterministic universe, we may still have free will. One of the appealing things about this theory is its suggestion that moral practices that depend upon free will could remain unchanged even if determinism were shown to be true. Compatibilism tells us, in other words, that determinism is not something we need to worry about—the discovery that determinism is true would not require a radical overhaul of many common-sense views about our moral lives. Thus, with respect to our moral practices, compatibilism is a relatively conservative and optimistic thesis.

Recently, however, Saul Smilansky has questioned compatibilism's claim to this conservative posture, arguing that the theory is in fact "a much more radical view than it is typically presented and perceived" (Smilansky, 2007, 347). As an example of this radicalism, Smilansky suggests that compatibilism does not have the theoretical resources to explain why we may not punish a person, whom we know will commit a crime, *before* she actually commits her crime. According to Smilansky, "compatibilism cannot resist in a principled way the temptation to prepunish people" (Smilansky, 2007, 347).<sup>1</sup> Since most people share the intuition that prepunishment is illegitimate, it may be that

---

<sup>1</sup> The debate in which Smilansky is engaged focuses largely on retributive punishment (as opposed to punishment that is justified by its value as a deterrent to criminal behavior). My discussion of punishment below should also be taken to refer to retributively justified punishment.

compatibilism has a very counter-intuitive consequence and that its complacency in the face of determinism is therefore unjustified.

Below, I sketch some objections to the practice of retributive prepunishment. Some of the objections I discuss turn on difficulties involved in making the kinds of predictions on which prepunishment would depend while other objections involve reference to general features of our views about the justification of criminal punishment. However, while these objections to prepunishment are consistent with compatibilism, compatibilism does not entail any of them. In other words, none of the objections to prepunishment that I discuss is a *characteristically compatibilist* objection to the practice. Since incompatibilists can rule out prepunishment on characteristically incompatibilist grounds, this marks an asymmetry between compatibilism and incompatibilism. Given our intuitive resistance to prepunishment, one might take this asymmetry to suggest that incompatibilism has a better claim than compatibilism to representing our common-sense intuitions about punishment. I suggest that this is not so: even granting our intuitive resistance to prepunishment, the fact that compatibilism cannot rule out prepunishment on internal theoretical grounds does not reveal it to be a particularly unintuitive or radical view.

### *1. Potential Objections to Prepunishment*

Saul Smilansky's recent discussion of prepunishment is related to an earlier debate between himself and Christopher New in which New argued that there is no absolute moral objection to punishing criminals prior to their crimes (New, 1992). Suppose we know that a person will intentionally commit a crime at a certain future time

and we also know that it will be impossible to punish this person after he commits the crime. New claimed that in some such cases it would be permissible to prepunish the would-be criminal and that it is only our inability to predict the behavior of others with precision that makes prepunishment generally illegitimate. In his original contribution to this debate, Smilansky argued against New that prepunishment is more than epistemically problematic, it is also morally objectionable because it is an affront to “the moral personality of the agent . . . who we must *respect as capable of not committing the offence*” (Smilansky, 1994, 52). More recently, however, Smilansky has argued that compatibilism, in particular, does not have access to this intuitive objection to prepunishment:

If we perfectly know *now* that it is completely determined that a person will commit a crime . . . the compatibilist does not have a strong principled objection to prepunishing this person now, before he has actually committed the crime. The common-sense objection, that we must allow him to change his mind, does not apply here; for according to compatibilism it is already determined that he will *not* change his mind (Smilansky, 2007, 348).

Can the compatibilist reject prepunishment? It must be admitted that it is no good for the compatibilist to insist that prepunishment is illegitimate simply because we happen to lack the ability to predict criminal behavior with the accuracy that prepunishment would require. Perhaps our inability to make such predictions will one day be overcome; thus, an objection to prepunishment that cites our epistemic limitations is defeasible. But one of the governing assumptions of Smilansky’s argument is that our intuitive opposition to prepunishment is not one that we are likely to overcome, so compatibilism shows itself to be counter-intuitive insofar as its opposition to prepunishment rests on merely contingent grounds.

There are, however, options available to the compatibilist who wishes to argue that prepunishment would be illegitimate even if we could predict criminal behavior with certainty. For example, a compatibilist might object to prepunishment for whatever reasons lead some people to regard retributive punishment as generally morally suspect. A person may believe that it is wrong to deliberately inflict suffering on others in the way that punishment usually does—and such a person may also be a compatibilist. While compatibilists believe that we may reasonably hold people morally responsible for their behavior if determinism is true, this does not entail anything about what views a compatibilist should hold on the general permissibility of punishment. Since it does not appear to be a commitment of compatibilism that retributive punishment is justified even in principle, a compatibilist is presumably free to regard all punishment, and thus prepunishment, as illegitimate.

Alternatively, a compatibilist who resists prepunishment may argue that the goals and social functions of retributive punishment are best achieved when it is applied *after* criminal activity. As with the preceding objection, this would provide the compatibilist with relatively fixed, theoretical grounds for rejecting a policy of prepunishment. According to Joel Feinberg's influential account, punishment often has an "expressive function" in our society: it is a "device for the expression of attitudes of resentment and indignation, and of judgments of disapproval and reprobation, on the part of the punishing authority himself or of those 'in whose name' the punishment is inflicted" (Feinberg, 1970, 98). On Feinberg's account, part of the function of punishment is to offer a morally significant response to criminal behavior and, more particularly, a response to the culpable sentiments and motives from which antisocial behavior issues. If

the point of punishment is to express our collective indignation at a criminal who is moved by antisocial sentiments, then perhaps this punishment will register with the criminal in the right way only if he has already been moved by the sentiments in questions. A corresponding point can be made from the perspective of institutions that administer punishment: if the aim of punishment is to make a move in a moral conversation with the one who is punished, then such a move may be most apt when the sentiments and actions that elicit punishment have already occurred.

However, the preceding analysis, whatever its merit, will not apply to every instance of punishment because not all punishments play the expressive role just discussed. Consider, for example, Christopher New's original case in which he suggests prepunishment would be an appealing option. In New's case, a police officer can preemptively issue a ticket to, and collect a fine from, a driver who plans to break the speed limit in such a way that he cannot be prevented from doing so and cannot be made to pay a fine afterwards (New, 1992, 35-36). A notable feature of this example is that it involves a criminal violation, and a corresponding sanction, that does not seem to fall within the sphere of Feinberg's theory. To put it in Feinberg's terms, speeding is a crime that is met with a "penalty" rather than a punishment; the difference is that penalties are not associated with the expressive significance of more serious punishments like incarceration (Feinberg, 1970, 98). Fines for speeding, or for parking in a restricted area, can be thought of as price tags associated with these activities rather than expressions of society's indignation and resentment.

If the expressive theory of punishment does not cover all the sanctions that a state might impose on its citizens, then perhaps we cannot appeal to this theory to explain why

prepunishment is illegitimate in New's example, but this is not necessarily a counter-intuitive result. If speeding tickets lack the moral significance and condemnatory force of more serious criminal sanctions, then perhaps common sense will see no reason not to issue a ticket in a case like New's in which society has no other way to receive the price it attaches to a certain activity. But in cases in which punishment *does* express social contempt and indignation for a criminal's motives, then prepunishment may not be the most felicitous means for conveying these responsive attitudes.

There is at least one more thing a compatibilist might say about why we should not prepunish. It may be that even if determinism is true (and we come to possess maximum crime-predicting power), we may still have an option that is superior to prepunishment. This superior option would be crime prevention.

Suppose that at some point in the future we develop a supercomputer capable of combining the deterministic laws of nature and the facts about the past to generate predictions about the times, places, and actors that will be involved in criminal activity.<sup>2</sup> In such a scenario, crime prevention would be possible in cases in which the predictions that the computer issues have a conditional form such as, "unless there is intervention, Smith will kill Jones at  $t$ ." By contrast, prevention would not be possible in those cases in which the supercomputer issues true categorical predictions like, "Smith will kill Jones at  $t$ ." If the latter prediction is true, then Smith will kill Jones at  $t$ , and that cannot be prevented.

There is reason to think that the supercomputer imagined here would regularly issue conditional predictions of the sort that would allow for crime prevention. At least

---

<sup>2</sup> The situation would be similar to the one described by Philip K. Dick in *The Minority Report*, though hopefully without the dystopian elements present in that story.

this would be so in cases in which publicizing the predictions would have a causal impact on the performance of the crimes in question. Suppose, for instance, that our supercomputer predicts that Smith will kill Jones and that this prediction is made available to law enforcement officers who, to all appearances, have the power to stop Smith from killing Jones: they know where Smith is, he is accessible to them in the normal way, he has not yet initiated a causal process that entails that Jones will die, and so forth. Now in a case like this, it may be impossible for the supercomputer to reach, and issue, a certainly true categorical prediction like “Smith will kill Jones at  $t$ .” The computer may not be able to confirm the truth of such a prediction prior to issuing it because in order to reach this conclusion the computer would have to calculate the effects of the true prediction on law enforcement. This means that the computer would have to come to a conclusion about the effects of its true prediction in order to reach its conclusion about what the true prediction is, so it would need to have its conclusion in order to reach it.<sup>3</sup> This would be an untenable position even for our supercomputer.

Of course, this problem would be avoided if we imagined our supercomputer to be churning out predictions in causal isolation from the subjects about which it is making predictions. But in this case, it is unclear how any prepunishments could be imposed. I suggest, then, that in many scenarios in which prepunishment would actually be an option, crime prevention would also be an option. And, even if prepunishment is a radical, counter-intuitive thing to advocate, crime prevention is not a radical innovation.

---

<sup>3</sup> This speculation is modeled on Hilary Bok’s discussion of the “Pocket Oracle,” a device rather like the supercomputer under consideration. “If . . . the Pocket Oracle tries to tell me in advance what I am going to do . . . it will run into problems. For in order to figure out what I am going to do, it must factor into its calculations all of the stimuli I receive . . . And if it plans to tell me what I am about to do or choose, the information it gives me will be among the things it has to take into account. . . . [it] must factor the result of its calculations into its calculations in order to arrive at a result; and it would have to know what that result was before arriving at it in order to do so” (Bok, 1998, 81).

Most people, I think, have little problem with the thought that it is sometimes justifiable to detain a person who clearly presents an imminent danger to herself or to others. One of the reasons such preventive steps seem reasonable is that they are not a form of punishment and need not be associated with, or motivated by, retributive impulses.

The foregoing sketches indicate, I think, several plausible routes a compatibilist might take to resisting a policy of prepunishment even in circumstances where we have maximum crime-predicting ability. However, the incompatibilist critic who raises the issue of prepunishment is not likely to be satisfied by any of the foregoing replies on behalf of the compatibilist. Smilansky, for instance, wants a *principled* objection to prepunishment from compatibilists and perhaps none of the above considerations would strike him as principled in the right way. What the incompatibilist critic presumably wants to hear is how it follows from distinctly compatibilist theoretical commitments that prepunishment in particular, and not punishment more generally, is wrong.

For her part, the incompatibilist can trace a direct argumentative route from her core incompatibilist commitments to a ban on prepunishment. She would say, for instance, that retributive punishment is only justified if the one who is punished is morally responsible for the behavior for which she is punished. But the sort of prepunishment discussed here relies on predictions that are made possible by the truth of causal determinism and since (according to the incompatibilist) determinism is incompatible with moral responsibility, prepunishment is illegitimate under these conditions.

Compatibilism cannot offer a corresponding—characteristically compatibilist—argument against prepunishment. The compatibilist should admit this much, I think, but



the question now is whether compatibilism is a radical thesis and at odds with common sense just because it cannot mount a distinctly compatibilist argument against prepunishment. I argue below that compatibilism is not shown to be a particularly radical thesis on these grounds.

## *2. What Does Common Sense Tell Us?*

Smilansky's advice to compatibilists is that they should "bite the bullet" and admit that the "only reason we ought not to prepunish . . . is epistemic, namely, that we rarely have the required powers of prediction" (Smilansky, 2007, 348). According to Smilansky, this shows compatibilism to be "in stark opposition to the common-sense view" about the wrongness of prepunishment (Smilansky, 2007, 348). This suggests that Smilansky takes our common-sense resistance to prepunishment to stem from something more than sensitivity to human epistemic limitations. But why should we believe that it is common sense, or a core moral intuition, that the wrongness of prepunishment rests on more than our inability to predict human behavior? Perhaps we have less reason to draw this conclusion than initially appears.

Smilansky says the "common-sense view" is that we should not prepunish because "we must let the (still innocent) person decide, even at the last moment, to refrain from committing the crime" (Smilansky, 2007, 348). I think that this is a fair characterization of common sense and that the substance of this claim is correct: we should wait until after "the last moment" before we hold someone to account for an action. However, it may be that what makes this a common-sense position is that waiting to the last moment is evidently the only sure way of finding out what someone will do.

But citing the fact that waiting until the last moment is the best way of finding out what will happen is just a way of characterizing one of our epistemic limitations. So perhaps the common-sense objection to prepunishment emerges out of sensitivity to human epistemic limitations after all. But in this case, compatibilism is not biting a bullet, or revealing a radical nature, when it objects to prepunishment on epistemic grounds.

If the common-sense objection to prepunishment turns on our epistemic limitations, then the compatibilist has access to this objection, at least under normal conditions. As far as the compatibilist need be concerned, it is perfectly true that we should not prepunish because, for all we know, any person we propose to prepunish may not commit the crime for which we punish him. What makes this hesitation to prepunish on the part of the compatibilist reasonable is that, as far as the compatibilist knows, determinism is false—and even if determinism were true, we still have nothing like the ability to predict people's behavior in the way prepunishment would require. At the beginning of the previous section, I quickly passed over the idea of objecting to prepunishment simply on the grounds of our inability to predict the behavior of others with certainty. Now it appears that this objection may have been dismissed too quickly. Perhaps it is perfectly consistent with common sense for the compatibilist simply to cite our epistemic limitations in explaining why prepunishment is wrong.

Of course, it is open to the incompatibilist to insist that the common-sense objection to prepunishment is *not* based on the fact that we are never sure exactly what the future holds. Instead, the incompatibilist might characterize the common-sense rejection of prepunishment as based on the view that punishable behavior must arise from an exercise of agency that is undetermined and intrinsically unpredictable—thus, if we

could predict an action with certainty, we would know that the behavior in question is not open to punishment. This is essentially the characteristically incompatibilist argument against prepunishment introduced above. Since the compatibilist will resist the view that determined behavior is *ipso facto* not punishable, this objection to prepunishment is not available to the compatibilist. But does this show compatibilism to be at odds with common sense? It would do so only if it were common sense that punishable behavior must emerge from unpredictable exercises of agency. But I know of no compelling reason to think that such a particular (and partisan) conception of agency and responsibility is a deliverance of common sense.

The situation, then, is this. Under real world conditions, characterized by familiar epistemic limitations, the compatibilist can adopt the common-sense perspective that we should not prepunish people because we cannot be sure whether a given person will commit a crime. However, we can imagine non-actual conditions in which determinism is true and we have the capacity to predict the future in such a way that prepunishment becomes a viable option. Under these conditions, the compatibilist would not have access to the common-sense objection just described. However, lack of access to this objection under these conditions is not a special feature of compatibilism. Under conditions of complete predictability, *no one* can plausibly say that our reason for not prepunishing is that we must give a person until the last moment to see what she will do; under the relevant conditions, the common-sense objection would fail for everyone, not only for compatibilists.

Of course, and as we have already seen, the incompatibilist still has an objection to prepunishment under conditions of determinism and predictability. This objection is

not that we must wait to see what a person does before we punish her, rather, it is that since we already *know* how the person will behave, the action in question cannot serve as a legitimate basis for punishment. This objection assumes that determinism and predictability are incompatible with the sort of agency required for justifiable punishment. Since the compatibilist obviously does not share these assumptions, this objection to prepunishment is not open to the compatibilist. But it is unfair to conclude that this shows compatibilism to conflict with common sense because it is unfair to the compatibilist to simply take for granted that common sense is characterized by incompatibilist assumptions. Now the incompatibilist might also mount an objection to prepunishment that does not rely on assuming core features of incompatibilism, but whatever these objections turn out to be, the compatibilist would presumably have as much access to them as the incompatibilist. Indeed, I offered objections to prepunishment above to which compatibilists and incompatibilists seem to have equal access.

### *3. How Radical is Prepunishment (Compared to the Alternatives)?*

Suppose that in the future we discover that determinism is true and we also discover how to use this fact to predict criminal behavior. Would it be radical for a compatibilist living in this future to view prepunishment as permissible? It would certainly be radical to think that prepunishment is permissible in the world as we know it. But this is not the perspective of our future compatibilist; he suggests, rather, that prepunishment is permissible in a world that is very different from the one we know because this future world is one in which we can accurately predict the future. Is this future compatibilist's view at odds with common sense?

The preceding question is difficult to answer partly because it asks us to judge the common-sense appeal of a practice that can arise only in circumstances that are very different from those that have conditioned our common-sense moral responses. Perhaps this should make us suspicious about any answer to the above question. Certainly, we should not simply assume that because there are good objections to prepunishment given the way the world actually is, it follows that we would, or should, find these objections natural in a world that is very different from the one we know.

But perhaps the incompatibilist will say that, regardless of whether we would find it sensible to prepunish in some science-fiction future, the fact that we are intuitively opposed to prepunishment at present means that it is common sense that prepunishment is forever and always illegitimate. In this case, if compatibilism can foresee circumstances where prepunishment would be legitimate, then compatibilism is at odds with contemporary common sense. But I do not believe that we should interpret common sense in the way just described. There is little reason, I think, to interpret our common-sense views on prepunishment as attempting to cover circumstances that are very different from those in which we are accustomed to finding ourselves. At least this seems to follow if, as I have suggested, the common-sense objection to prepunishment is largely motivated by awareness of human epistemic limitations. Since these limitations may prove to be contingent, I suggest that the common-sense prohibition of prepunishment is a contingent prohibition. The preceding line of thought involves two related claims: (i) while it would be radical to advocate prepunishment in the world as we know it, it is not clear that it would be radical to advocate prepunishment in a future in which we possess the requisite predictive capacity; (ii) it is not clear that it is now contrary to common

sense to think that there may be a future in which prepunishment would be permissible in some cases.

It is worth considering the alternatives to prepunishment in our hypothetical future of determinism and complete predictability. As I noted above, the idea that we must give a potential wrongdoer until “the last moment” to avoid wrongdoing and punishment loses its force in this futuristic context. Let us suppose that without access to this common-sense objection to prepunishment, future compatibilists decide that prepunishment is legitimate in some cases. For their part, future incompatibilists may still object to prepunishment because they take determinism and predictability to be incompatible with the sort of moral responsibility required for justified punishment. But note that this incompatibilist response is an objection to normal “postpunishment” as much as it is an objection to prepunishment. In the future world we are considering, the incompatibilist’s only option to permitting prepunishment would be to view *all* punishment as illegitimate. But perhaps opting for prepunishment is no more a radical alteration of our present practices than would be abandoning punishment altogether. So, again, opposition to prepunishment under conditions of determinism and predictability may have no greater claim to representing common sense than would support of prepunishment under those conditions.

As I argued in Section 1, a compatibilist could make use of certain objections to prepunishment even under conditions of determinism and predictability. But while the considerations that ground these objections are consistent with compatibilism, they are not entailed by the basic theoretical commitments of compatibilism. The incompatibilist may therefore argue that because no objection to prepunishment follows from the basic

tenets of compatibilism, the view by itself does not exclude the permissibility of prepunishment. In this case, compatibilism, left only to its basic theoretical commitments, cannot predict confidently that determinism would give us no reason to alter our moral practices.

Perhaps it follows that compatibilists somewhat overstate the case for optimism in the face of determinism; perhaps compatibilism cannot entirely exclude the possibility that determinism will change some of our practices. Smilansky may be correct, then, when he says that “[t]he traditional compatibilist stance, according to which determinism does not really change anything, morally, is . . . *shown* to be false” (Smilansky, 2007, 348). However, I do not think that Smilansky is right to conclude from this last claim that “[t]hinking about prepunishment reveals what a radical position compatibilism is” (Smilansky, 2007, 348).

To see why Smilansky is wrong to draw this conclusion, it helps to keep in mind that the truth of determinism alone would not convince a compatibilist to prepunish—the truth of determinism would have to be coupled with possession of the remarkable ability to predict the future. Indeed, much of the strangeness in the idea of prepunishment may lay not so much in the practice itself as in the thought that we might possess the predictive power that would make the practice feasible. The very idea of having this power is strange and disconcerting, but the idea that such a fantastic power might alter our practices is, I think, much less odd. One might even say that it is a common-sense expectation that a radical change in how we acquire knowledge about the future would have some impact on our social institutions.

So perhaps the compatibilist should admit that she cannot rule out the possibility that discovering the truth of determinism would lead us to make some changes in our moral practices. But, at least with respect to the issue of prepunishment, I think the compatibilist can claim that the truth of determinism need not lead us to change our practices in a way that would be in obvious conflict with our moral intuitions. If a future compatibilist were to judge that prepunishment is justified in some cases, I suspect that these would be cases like the one New described in which we *know* both that an unstoppable legal violation will occur and that we can impose a socially useful penalty prior to the violation, but not afterwards. In a case like this, there seems to be good reasons to impose the penalty prior to the violation. Quite possibly, imposition of the penalty would seem like common sense under these circumstances. Of course, many future cases of wrongdoing would be dissimilar from New's example in one way or another. I suspect that the compatibilist would have access to the sorts of considerations discussed in Section 1 to explain why prepunishment is best avoided in those cases.

### *Conclusion*

It would be useful for the incompatibilist to be able to argue that compatibilism has unintuitive consequences and that it cannot live up to its reputation as an optimistic view that promises to preserve our moral practices in the face of determinism. Perhaps this sort of argument can be made, but I do not think it will turn on the issue of prepunishment. First, there are certain general reservations about prepunishment that may be held by a compatibilist as much as by anyone else. There may be no characteristically compatibilist objection to prepunishment in the same way that there is a characteristically



incompatibilist objection to the practice, but it is not clear that this puts compatibilism at odds with common sense since it is not clear that our intuitive opposition to prepunishment is founded on deep-seated incompatibilist assumptions.

It may well be that the widespread, common-sense resistance to prepunishment hinges on the fact that waiting until people act is, as it happens, the surest way to find out how people will behave. If this is so, then compatibilism has access to the common-sense objection to prepunishment under real world conditions. There may come a time when this objection to prepunishment becomes inoperative because we have some other way of knowing how people will behave. This would be a radically new way of acquiring knowledge and under these circumstances—and in certain restricted cases—prepunishment might well commend itself to common sense. At any rate, a world in which we can scientifically predict human behavior is sufficiently different from the world we know that we should be wary of making definitive pronouncements (one way or the other) about what common sense would tell us under such conditions.<sup>4</sup>

---

<sup>4</sup> I would like to thank John Martin Fischer, Michael McKenna, Neal Tognazzini, and Saul Smilansky for their helpful comments on an earlier draft of this paper.

## REFERENCES

- Bok, Hilary. 1998. *Freedom and Responsibility*. Princeton University Press.
- Feinberg, Joel. 1970. "The Expressive Function of Punishment." Reprinted in Feinberg, *Doing and Deserving: Essays in the Theory of Responsibility*. Princeton University Press.
- New, Christopher. 1992. "Time and Punishment." *Analysis* 52: 35-40.
- Smilansky, Saul. 1994. "The Time to Punish." *Analysis* 54: 50-53.
- Smilansky, Saul. 2007. "Determinism and Prepunishment: The Radical Nature of Compatibilism." *Analysis* 67: 347-49.