Pereboom on Punishment: Funishment, Innocence, Motivation, and Other Difficulties

Saul Smilansky[1]

Abstract In *Free Will, Agency, and Meaning in Life*, Derk Pereboom proposes an optimistic model of life that follows on the rejection of both libertarian and compatibilist beliefs in free will, moral responsibility, and desert. I criticize his views, focusing on punishment. Pereboom responds to my earlier argument that hard determinism must seek to revise the practice of punishment in the direction of *funishment*, whereby the incarcerated are very generously compensated for the deprivations of incarceration. I claimed that funishment is a practical *reductio*: of hard determinism. Pereboom replies, but I claim that he misses a key component of my *reductio*, the idea that moving in the direction of funishment will considerably weaken the deterrence of potential criminals so that hard determinism becomes self-defeating in practice. Beyond the challenge of funishment, I raise various other difficulties with Pereboom's model, concerning its deeply unintuitive implications, the harm it does to the motivation of potential criminals, its weakness in resisting utilitarian-like dangers, and more. Our conclusions should lead to a re-evaluation of the compatibilist interpretation of moral life, as a richer, more plausible, and safer interpretation than hard determinism. This needs to be combined with a true hard determinist acknowledgment of the deep injustice and tragedy involved in punishment in light of the absence of libertarian free will. Such a complex view will come closer to doing justice to notions of justice, morality, and decency.

Keywords Pereboom · Smilansky · Free will · Moral responsibility · Desert · Punishment · Funishment · Innocence · Motivation

In *Living without Free Will* (Pereboom 2001), Derk Pereboom sought to draw a picture of moral, social, and personal life without belief in free will and moral responsibility. *Free*

✉ Saul Smilansky
  smilsaul@research.haifa.ac.il

1   Department of Philosophy, University of Haifa, Haifa, Israel

 Springer

Will, Agency and Meaning in Life continues this project, responding to criticism and developing the model, in his usual ingenious and systematic ways. The picture Pereboom offers is distinctly optimistic: living without free will and moral responsibility of either the libertarian or compatibilist varieties is, at worst, almost as good and often (as with punishment) better. In his view, we have no free will and moral responsibility but this is no cause for mourning, and indeed it is often cause for celebration—we are, more often than not, better off without it. The great optimism therefore leads to an eagerness to change our moral, social and personal beliefs, attitudes, and practices in a way that would reflect the absence of free will and moral responsibility.

My view is very different (Smilansky 2000, 2005). I share Pereboom's rejection of libertarian free will (LFW) but beyond that our ways diverge. I believe that compatibilism is partly true, and that we can and should learn from its insights. There are viable forms of compatibilist free will, and these allow for a Community of Responsibility, and moral and personal beliefs, reactions, and practices, that reflect compatibilist distinctions and differences in the degree of compatibilist freedom and responsibility. In some contexts, the compatibilist interpretation of the implications of living without LFW is more salient, and compatibilist free will suffices. However, compatibilism often comes up short, and the absence of LFW greatly matters, e.g., making the best compatibilist reactions and practices deeply unjust or limited. A true understanding of the free will problem is hence fundamentally complex, and the most adequate position on the compatibility position is thus a "compatibility-dualism." Moreover, the absence of LFW and the severe limitations of the compatibilist substitutes are by and large negative. Living under hard determinism (HD), as Pereboom proposes, is in many ways impossible, and even where it is feasible, it is highly undesirable. This pessimism thus leads me to conservatism on free will. At the edge, it pushes me towards Illusionism: the idea that prevalent false beliefs (primarily in LFW) are by and large positive, and should not be tampered with outside of philosophy. Illusion helps to maintain the true but limited and vulnerable compatibilist beliefs, and also guards us against the dangers of a fuller realization of the tragic moral, social, and personal implications of the absence of LFW.

In this essay, I will focus upon Pereboom's views on punishment. First, I will continue our debate on my "funishment" argument (Smilansky 2011; Pereboom 2014). This is a practical *reductio* argument against "happy hard determinists" such as Pereboom. (When I speak about HD, I refer to views which reject robust forms of LFW, either because of determinism or because of indeterminism in that e.g., Quantum Mechanics does not help. I prefer to use the traditional terminology over Pereboom's "hard incompatibilism.") Second, I will note more briefly some other difficulties of HD in the context of punishment. I will try to show how HD is too *thin a*nd does not have the resources to address the needs of real life and the concerns with justice and decency.

The free will paradigm holds that free will is a necessary condition for moral responsibility, which is a condition for desert, which is required for just punishment. Libertarians and compatibilists differ, of course, in the way they understand the needed form of free will but both can justify punishment within the paradigm. Both can find punishment morally acceptable as it is commonly practiced, involving long-term incarceration in unpleasant institutions set apart from lawful society. This need not even involve the idea that punishment is intrinsically good. If such punishment is socially useful (for purposes of deterrence, for example), libertarians and compatibilists may not object to it, since they would accept that the criminal has made himself liable to be disposed of in harsh ways, through his free actions. By contrast, hard determinists believe that there is no free will,

and so no moral responsibility, and no desert. How, then, can hard determinists deal with the need to punish, when coupled with the obligation to be just?

I have argued that hard determinists are in fact committed to *funishment*, instead of punishment. But, by its nature, funishment is a practical *reductio* of HD: it makes implementing HD impossible to contemplate. Indeed, the social practices that HD requires turn out to be morally bad even according to HD itself. What is funishment, and why do I claim that HD is committed to it?

1 My Practical *Reductio* Argument Against Hard Determinism

I will summarize this argument (Smilansky 2011) in 10 stages:

1. Murderers, rapists, violent bullies, thieves, and other miscreants need to be kept apart from lawful society; we have no real choice, for otherwise they will kill, rape, steal, and make life miserable for the rest of us.
2. Of course we should aim to reform social conditions that lead many to crime. Also, nothing in my argument requires that we accept that current levels of punishment are optimal, and in fact it is likely that we over-criminalize and over-punish. The ways in which we punish are also in need of significant reform; we should for example make greater efforts to rehabilitate. But, even after we take into account all plausible possible changes and reforms, a large amount of long-term criminal incarceration seems unavoidable, given realistic social prospects and the human condition as we know it. Even the most humane societies have not found a systematic substitute for incarceration, as a way of deterring potential offenders and keeping actual offenders off the streets.
3. Since HD holds that no one deserves the hardship of being separated from regular society, this hardship needs to be *compensated* for. Hence a great effort and expense must be made, in order that a person undergoing punishment will have a good life despite being separated from regular society, and deprived of the freedom and opportunity to move among the rest of us.
4. This effort at compensation morally must receive high social priority. We are at risk of ruining the lives of people who (again, assuming HD) in no way deserve this. Incarceration does not just occur. The criminals are proposed to become *our* victims: they will suffer grave deprivation and harm (in itself, and relatively to others on the outside), due to our own intentional actions, and because this serves our interests. We, as a society, are proposing to target and injure them in a special way, which they do not deserve, for our own purposes. In order to be permitted to do so, we must provide adequate compensation.
5. So, instead of punishment, we should have *funishment*: funishment would resemble punishment in that criminals would be incarcerated apart from lawful society; and institutions of funishment would also need to be as secure as current prisons to prevent criminals from escaping. But institutions of funishment would also need to be as delightful as possible. They would need to resemble five-star hotels, where the residents are given every opportunity to enjoy life.
6. This would go beyond material conditions: each criminal will need to be permitted considerable leeway in running his or her own personal life, as well as a large measure of freedom of social interaction (including frequent visits from outsiders).

7. Criminals currently have to balance the temptations of crime with the risks of punishment: the risk that, if caught, they are likely to spend many of the best years of their lives in miserable, ugly, harsh, nasty, violent, and otherwise highly unpleasant institutions. Some people nevertheless take the risk, while many others are deterred. But, once funishment replaces punishment, matters change radically. The potential offender knows that, if he is not caught, he can enjoy the spoils of his crime. But even if he *is* caught, he faces only some time in an institution of funishment, which—apart from being separated from lawful society—will be like a fabulous holiday. Funishment will greatly weaken the deterrence of institutions of punishment.

8. My claim that a system which could threaten potential offenders with, at worse, funishment, would be challenged in its efforts to deter, while empirical, can hardly be controversial. Modern societies are finding it difficult to deter many people even with present, highly unpleasant prisons; a turn towards funishment must greatly weaken deterrence. Following HD would lead to a flood of crime. The number of people who would need to be kept apart from lawful society would increase enormously. Many people who would otherwise not have become involved in crime, nor ever suffer detention, would be caught up in that very life. In the meantime, the rest of us would be living in the worst possible world: suffering unprecedented crime waves while paying unimaginable sums for the upkeep of offenders in opulent institutions of funishment.

9. Even in hard determinist terms, all this is a very bad state of affairs. Hard determinists have sought to limit the number of people with which the justice system must deal, to reduce public hatred of offenders, and to beneficially reform the social conditions that generate crime. But HD itself defeats all those idealistic goals. If implemented, the view would generate more rather than less crime, more criminals would be caught up in the system and incarcerated apart from society (albeit under improved conditions), and public sentiment would hardly move towards an offender-sympathetic stance, once crime blossoms, and the taxation required to finance the regime of funishment mushrooms. This makes a backlash against *funishment* very likely.

10. In any case, a hard determinist order would be nightmarish, even for HD, if *correctly* implemented. Hard determinists themselves cannot desire the results of the reforms required by their own position (rising crime, much higher levels of incarceration, etc.). HD is, in practice, *self-defeating*.

2 Pereboom's Response

Pereboom responds to my argument (Pereboom 2014: 172–173). He first says that "The details of the skeptic's reply to Smilansky's objection depend on which general theory she thinks can be defended. My own predilection is to endorse an axiological moral theory which includes better consequences as valuable, where morally fundamental rights being honored and not violated count among good consequences" (p. 172).

Pereboom then admits that "the preventively detained would not be treated worse than needed against the danger they pose. In addition, the right to live a fulfilling life is in play and weighs heavily, and we would thus have a serious moral interest in providing those who are preventively detained with requisite opportunities and conditions. On the issue of

cost, providing these sort of opportunities may add expense to our system for dealing with criminal behavior, but not the expense required to provide all of those detained with 'five-star hotel' accommodation" (p. 173).

However, I cannot see how Pereboom meets my challenge. He focuses on the issue of expense. He himself admits as we have just seen that institutions of detention may "add expense to our system." But the expense is only one of the ways in which, I claim, implementing HD in the sphere of punishment would be counterproductive, even in terms of HD (of the sort he favors), and thus leads to a *reductio*. The main issue is that of *deterrence* but, as far as I can see, Pereboom does not address it.

Consider step #6 of my argument, above: "This would go beyond material conditions: each criminal will need to be permitted considerable leeway in running his or her own personal life, as well as a large measure of freedom of social interaction (including frequent visits from outsiders)." Allowing this, in what also needs to be a secure prison, would surely add considerable expense. But even setting the expense aside, it would *crucially weaken the deterrence of punishment*. Admittedly a few prisoners could not be trusted to confront anyone, including their families and friends, but the vast majority of criminals could. Knowing that, if you were caught, your family and friends could visit you, and indeed stay with you for some time, would take away a major consideration currently deterring most people from criminal activity.

Pereboom likens punishment to quarantine, which is justified by the dangers of allowing those quarantined to engage freely with others (more on this below). He insists that the conditions of the punished/quarantined ought not to be more severe than necessary, given the danger they pose. But typical criminals are not contagious; there is no reason, e.g., why their girlfriends or boyfriends, or spouses, could not come to live with them. The quarantined would stay in prison, and their friends or spouses could go in and out, and continue in daily life. But, again, such arrangements—which I do not see how Pereboom could resist—would only further weaken the deterrent.

The question of expense also does not depend on exactly where we put the level of comfort of the prisoners. While I think that the evil of deprivation of freedom requires compensation at the equivalence of a five-star existence, perhaps Pereboom thinks only two-stars are needed (he admits that additional expenses over those of today would be needed). But the crucial point is that even a two-star existence, with the freedom to receive guests and all that would be involved in the radically-revised prison situation we are considering, would weaken the deterrent effect. And then, even if the increase in expenses pertaining to current prisoners is not overwhelming, *when we add the number of additional new prisoners led to crime by the weakened deterrent motivation*, this will become a major financial consideration. Of course it is not only a financial matter—even if we bracket the issue of added expense. HD defeats itself, if implementing it leads to a massive increase in criminals, and hence in the numbers that will be incarcerated.

Libertarians and compatibilists have the resources to reject such obligations to funish (rather than to punish), since the criminal deserves the limitations and hardship due to her free actions. She brought them upon herself. But hard determinists do not: HD is a morally much thinner theory. Losing free will and moral responsibility-based justification hence makes HD vulnerable to my *charge of the disappearing deterrence*, thus leading to the *reductio* of this position. Even in its own terms, HD is unworkable.

In the foreseeable future, even hard determinists will need to incarcerate criminals away from lawful society. They cannot, however, accept common forms of punishment, in the way that both libertarians and compatibilists can. I have argued that morally serious hard

determinists are committed to move in the direction of funishment instead of punishment, and that this creates a practical *reductio* of their position.

3 Where Does this Leave Hard Determinism?

If my funishment *reductio* works, then Pereboom's optimism on the benefits of abandoning belief in free will and moral responsibility becomes unconvincing. He is left with various options, all of them rather grim. He could, for example, simply opt for a utilitarian-like consequentialism. Pereboom favorably quotes Neil Levy, from the latter's response to my paper: "A consequentialist who is a moral responsibility skeptic will naturally hold that no one should be treated any worse than is needed to bring about the best consequences, with all agents' welfare—including the welfare of criminals—taken into account" (Levy 2012). But it seems that Pereboom's over-optimism does not allow him to see that he is at great danger here. Levy is a (broadly, utilitarian-like) consequentialist, and hence is presumably ready to do whatever it takes to prevent most of the harms of severe crime (albeit considering also the welfare of criminals).

Pereboom, by contrast, repeatedly warns against using a person "as a means" (e.g., p. 165), speaks about fairness as a problem of utilitarianism (e.g., p. 164), and in other ways seeks to constrain a general utilitarian-like consequentialism. If my argument succeeds, he could not continue to have his cake and eat it too. The needs of deterrence would make any effective incarceration nearly as nasty as current prisons. HD would need to raise up its hands and surrender, or embrace a utilitarian-like consequentialism.

Moreover, if we are thrown at the mercies of such a theory, then the traditional worries about utilitarianism emerge. Utilitarianism has, of course, been available for a long time, but people who worry about free will and moral responsibility are concerned with more than utility. This concern has been the basis for taking the moral high ground as against a solely utilitarian justification of blame, guilt, and harsh punishment, rightly saying that making people suffer guilt or punishment *just because doing so would be socially useful* is morally unacceptable. We cannot use the morally innocent in such ways even if it furthers social interests. Blame and punishment must be just, and not only socially efficient. And, in order for them to be just, they must follow upon the choices and actions of moral agents who, through their free actions, have made themselves liable to blame and punishment. That is why the punishment of the innocent is the paradigm of injustice (yet it is of no principled concern to a utilitarian). Pereboom shares an intuitive dislike of utilitarianism because of the punishing of the innocent issue, saying, for example, that "this practice [a scheme of punishing the innocent] would seem to be more deeply wrong than can be accounted for by the utilitarian reasons…" (p. 165). But the difficulty is that, *for hard determinists, everyone is morally innocent!* HD as a moral position thus ought to hold that no one deserves to be made to suffer, or to be made worse off than another, and hence that it would be *unjust* to do so. Pereboom ought to acknowledge that this is indeed an injustice. HD cannot turn to consequentialism for assistance in overcoming the *reductio*, for it would thereby completely betray itself as a distinct ethical position.

The terminology can mislead us here. HD could simply be taken to be the conjunction of determinism with incompatibilism, namely, as the view that determinism exists, while the sort of ("compatibilist") free will that could ground moral responsibility in a deterministic world does not exist. This understanding would be adequate for certain purposes, but not in the present context. For it leaves it an open question whether we may punish

people (say, based solely on utilitarian-like criteria). In our context, such a narrow understanding of HD makes it completely mysterious as to why we care about the free will problem in the first place. HD understood as a deep moral view shares the broadly deontological paradigm notion of free will (the idea that free will is required for moral responsibility, which is a necessary basis for desert and punishment) with libertarianism and compatibilism. But, unlike them, HD is pessimistic as to whether the right sort of free will can exist. This pessimism would not suddenly make a morally serious hard determinist indifferent to the fact that we do not have free will and moral responsibility, and permissive about blame and punishment. On the contrary, such a hard determinist thinks that *because* we lack free will and moral responsibility, it would be *morally wrong* to punish (or blame, etc.) anyone.

So, as far as I can see, if my *reductio* argument works, Pereboom must either sail out to sea with a dangerous utilitarian-like consequentialism, and justify a fairly nasty and unjust form of incarceration in order to maintain the deterrent effect of punishment, or declare the practical bankruptcy of HD in the sphere of punishment. HD hence confronts a vicious philosophical fork: attempting to maintain its integrity leads it to failure in practice and to self-defeat even on its own terms; while attempting to avoid those grim outcomes through embracing a utilitarian-like consequentialism leads it to moral and philosophical self-betrayal. In either case, this should take the "happiness" out of "happy HD," and lead Pereboom to admit the dominance of injustice and tragedy even in the best realistic schemes of punishment.

4 Innocence, Constraints, Motivation, and Other Difficulties

So far I have concentrated on the implications of my "funishment" argument, which show that HD could not be implemented, because of the weakening of deterrence which would follow from the need to turn punishment in the direction of funishment. Now I would like to move in other directions, and explore further difficulties of following in the direction Pereboom recommends, focusing on the questions of *who* will be incarcerated, for *how long*, and what safeguards will be provided during incarceration. In the process, we will see other reasons for concern about the impact of adopting HD, such as that regarding the satisfaction of the families of the victims of crime, and the effect on the motivation of some potential criminals.

For Pereboom, the model for punishment is the quarantine: "Suppose someone poses a danger by threatening to commit murder. Even if he is not in general a morally responsible agent in the basic desert sense, the state would nevertheless seem to have as much right to isolate him as it does to quarantine a carrier of a deadly communicable disease who is not responsible in this sense for being a carrier" (p. 169).

In my view, the punishment-as-quarantine model raises grave difficulties. First, the analogy of the quarantine is of only limited help. Quarantine is, in modern times, quite rare, is typically limited in duration, and is mostly imposed on foreigners. Our intuitions on it are thus not well formed in a way that would be informative in the present context.

Second, quarantine is thought to be justified only in cases of immediate danger of the gravest sort. Those features are not present in most cases when we want to imprison people. Most criminals are imprisoned for crimes that are not life threatening, and so could not be quarantined at all, according to the standards we demand for medical quarantine. Note how reluctant we are to quarantine AIDS sufferers, despite the potential dangers of

contagion (and this was so even at the time when catching AIDS would have been fatal). This is already a difficulty.

But let us see what happens if we take the plunge, and explore punishment-as-quarantine. Using commonsense, some people *obviously* ought to be punished: say, an otherwise law-abiding and moral man, who quite reasonably hates his ill wife (who abused him throughout their marriage), deprives her of her medication in order to shorten her life. Or, children who tamper with a medical device in order to hasten the death of a very wealthy but obnoxious parent from whom they will inherit. Yet the quarantine model may well not yield the required result. All the man wanted was to get rid of his wife, and there is no reason to think that he will pose any social danger once this unique problem is solved for him. Likewise, the children have always been normatively well behaved, and will continue to be so; it was just extreme financial pressures that led them to do it, and those pressures have now been permanently relieved through the inheritance.

Libertarians and compatibilists will, of course, be able to justify punishment of these wrongdoers through their models, but a hard determinist such as Pereboom, who is committed to quarantining only those likely to pose a danger in the future, seems not to have the requisite philosophical resources here. The question is not only practical; HD simply seems to give the wrong moral answer. But surely it also puts HD in an extremely awkward practical situation: try to explain the thought that the man who murdered his wife need not even be quarantined (let alone more traditionally punished) to her family, for example. And, while it might reasonably be thought that actual cases of this sort are not very common, surely much of this results from the deterrence of traditional free will-based notions of punishment. The temptations of hatred and inheritance are there for many people, and once it becomes known that "one-timers" will be unharmed, one-time crimes will become even more tempting.

The predicament of the relatives of the murdered woman can help us to see another difficulty with Pereboom's model of punishment: the issue of public perception. Hard determinists have often emphasized the idea that punishment as commonly carried out fulfills personal and public vengefulness and cruelty (Bruce Waller, another "happy hard determinist," has made this claim in detail; see Waller 2011). But the above examples show that the worry does not depend upon unreasonable vindictiveness. Similarly, we can see that the HD obligation to "funish" (even if not in the opulent form I think is required by the need for compensation for the incarcerated) can be expected to have a highly disturbing impact. State punishment needs to substitute for the natural desire of relatives and friends of those wronged to get back at the wrongdoers; but funishment, let alone the setting free of those who do not (or no longer) pose a threat, is very problematic here.

Pereboom might reply that the public needs to be educated and that, when more and more people accept HD, they will acquiesce in funishment, or even in the absence of any punishment. At the very least, HD has yet another grave practical problem here. The public is only marginally ready to acquiesce in the hospitalization of mentally ill criminals, and objects to their quick release even if cured. The idea that many criminals will not be incarcerated (because their wrongdoing is likely to be a one-time affair; or they are otherwise non-threatening), and that those who will be incarcerated will live under conditions of funishment, is unlikely to be thought morally acceptable. This weighs in the balance against the acceptance of the hard determinist interpretation of the moral situation, given the absence of LFW (more on this below). And, in any case, it creates further obstacles to the practical workability of HD of the sort that Pereboom favors.

No less worrisome is the potential effect on criminals. Consider a man contemplating violently attacking some innocent people at their homes and robbing them. Earlier, I asked

about the expected difference in his motivation, were he to think himself risking common punishment versus funishment. Now let us hold the expected conditions of incarceration constant (well-short of funishment). In the present thought experiment, we will consider the state of mind of the given criminal, in terms of his beliefs about the way his actions will be considered, in free will-relevant ways. If he himself believes in free will and moral responsibility, then he is more likely to have compunction concerning his actions, which, after all, he holds himself free and responsible in doing. Moreover, when he contemplates the views and attitudes that others will have towards him, he is more likely to hesitate if he thinks that others will hold him responsible, blameworthy, despicable, and deserving of punishment. Hard determinist beliefs are likely to have the opposite effect. This sort of consideration is related to what I named (Smilansky 2000: 153) "The Present Danger of the Future Retrospective Excuse," namely, the negative effect on our motivation to behave well, of the thought that even if we do not, then in the future, when our present actions are retrospectively reviewed, we will have a ready (free will-based) excuse.

Admittedly the views of criminals are likely to vary, and would sometimes be perverse. But, as with previous points, traditional free will beliefs have some motivational resources that are simply unavailable to hard determinists.

If HD of the sort that Pereboom favors seems to give a too-lax answer in some cases (e.g., not incarcerating one-timers), it is likely to be too severe in others, giving recommendations that would be morally disturbing and indeed frightening. I need not spend much time showing this, and can refer to familiar arguments posed against the "medical" alternative to the traditional criminal justice system. The idea that wrongdoers are merely ill, and their wrongdoing needs to be treated (rather than themselves being blamed and punished), may seem progressive and humane. But, as has been often argued (e.g., Hart 1970; Murphy 1973; Morris 1976), this is too quick. Free will-based models of punishment have the resources to limit punishment to actual wrongdoers (bracketing epistemic issues). They are also typically constrained in their punishment by the proportionate moral "price" the given crime merits, so that one will be punished only for and to the degree of the given crime that one committed. And there are also severe limitations on the permission of breaching the integrity and autonomy of the punished.

Medical models are not so constrained. Since the main concern is the forward-looking "healthy" social functioning of the potentially incarcerated, they could be deprived of their freedom even if they have not done anything. They may be deprived of it for a length of time that is totally disconnected from anything they have done: unlike the free will-based model, with the medical model, once one is in, there is no knowing in advance as to when one will be released. And, finally, since the concern is with changing the ways of the incarcerated, constraints over behavioral and medical treatments are much weaker. All this pertains to legitimate practices; the medical model also raises considerable opportunities for "Cuckoo-nest-like" abuse. One need not be blind to the wrongs and horrors of the current justice system to see that the medical model poses special wrongs and risks.

Another danger is more "statistical," and concerns the question of how "fussy" we will be about innocence. Utilitarianism has traditionally come in for criticism of this sort, as we have already noted. But most such criticism has focused on extreme examples, such as scapegoating. As I have shown long ago (Smilansky 1990, 2000: 28–30 with relevance to free will; cf. an excellent development of my argument in criticism of the quarantine model by Lemos, unpublished), the real danger here concerns a moderate lowering of standards for prosecution and conviction. On utilitarian grounds, it typically makes no sense to intentionally target an innocent person; and even less so to devise a system for doing so. This is too risky, and likely to be counterproductive. But the risks we take of punishing the

innocent versus letting criminals go in standards of evidence, for example, depend, for utilitarians or their kin, on the circumstances and potential risks and benefits. Our current practices put a premium on avoiding the risk of punishing the innocent, even when we are well aware that we thereby permit criminals to continue to harm innocent victims. Obviously any system will need to create a balance here, but the point is that those concerned with maximizing expected utility rather than with desert will put it in a place that will increase the risk of the punishment of the innocent.

Pereboom is aware of these issues, and seeks to address them by giving a special axiological weight to the rights of the incarcerated (p. 172). But the problem is, once again, that as a hard determinist, his moral-philosophical resources are much weaker than those of his free will rivals, and in exactly the places that matter. A free will-dependent justification for punishment can resist the medical model, and give replies as to who will be incarcerated, for how long, and how she may be meddled with, that are consonant with moral common-sense. Such constraints are not available to a hard determinist. Similarly, the utilitarian-like temptations for making the system of punishment efficient can be resisted much more forcefully if we take seriously the distinction between the morally guilty and the morally innocent. But, since hard determinists reject free will, moral responsibility, and desert, and do not give them inherent moral weight, *they are in a much weaker position to resist practices that are socially efficient but increase the chances that the innocent (in traditional terms) will be wrongly prosecuted or convicted*. Pereboom's hopes and good intentions are simply not supported by the resources of his view on free will and, once again, the loss involved in the rejection of free will and moral responsibility looms large.

Moral-historical progress has been exhibited in considerable measure in the rise of beliefs in free will and moral responsibility, in their importance for autonomy and individual empowerment, and in the rejection of ideas of blame and punishment that are not related to individual choice and responsibility (as in collective punishment). History shows that such progress is fragile and can be reversed (cf. Smilansky 2001; Sommers 2012). The calls for efficiency often suffice to sidetrack justice and decency. A similar story emerges if we compare criminal with distributive justice: distributive justice is not taken very seriously when confronted by pragmatic considerations. We are, for example, ready to acquiesce in not pursuing what is considered just taxation, in cases where doing so will lead to a decrease in revenues or otherwise reduce the economic pie. Justice can be neglected all too quickly, and eliminating free will-based support is a dangerous thing.

In fact, even a utilitarian-like consequentialist view is likely to oppose the widespread acknowledgment of HD, even if it were the true view. Its practical and moral limitations are too great, it is impractical, and it has many negative effects, as we have seen. To the extent that we are not such monistic consequentialists (and neither Pereboom nor I are), the difficulties of HD in resisting such consequentialist engagements, which we have seen in detail, should be added to the balance.

5 Hard Determinism or Compatibilism?

Again and again, we have seen the weaknesses of HD, in meeting the practical demands of real life, and our basic moral intuitions about justice, rights, and decency. This cannot plausibly affect the likelihood that a robust form of LFW is, after all, true. LFW is a straightforward view of the world, e.g., about the coherence and likelihood of the existence of agent-causality. On the other hand, this systematic failure of HD should lead to a re-

evaluation of its rival, compatibilism. HD and compatibilism ("soft determinism") are *interpretations* of determinism or of the implications of the absence of a robust LFW irrespective of determinism. Since they are both only interpretations, there is room for re-considering the rival interpretation, in view of the grave philosophical, moral and practical price that HD incurred. Hard determinists such as Pereboom set the bar for free will and moral responsibility very high, and since compatibilist notions cannot jump it, declare against free will and moral responsibility.

However, our conclusions should suggest the thought that perhaps the bar has been set too high. Repeatedly we have seen the limitations of HD, as compared to compatibilism—in resisting funishment, getting the punishment of one-timers correct, resisting the punishment of the innocent (in the sense of those who have done no wrong), maintaining fairness and decency in the face of the temptations of the medical model, meeting the call for justice of the family and friends of the victims of crime, and buttressing the moral motivation of potential criminals. Repeatedly, HD was seen to be too thin, a moral model that simply could not do the work on punishment, practically and morally. In my opinion, Pereboom should consider lowering the bar, thinking of desert not in terms of "basic desert" but in more moderate ways, and opening the possibility for acknowledging free will and moral responsibility, in degrees, according to compatibilist distinctions. The issue of punishment is too limited to be conclusive, but a similar case can be made, in my view, concerning other matters in moral, social, and personal life. Much of the actual way in which Pereboom engages with the world resembles revisionist compatibilism such as that of Vargas (2005, 2013). Instead of positing free will and moral responsibility impossibly high, Vargas proposes a forward-looking model that is in many ways similar to Pereboom's, but avoids the philosophical and ethical impoverishment to which hard determinists are led by interpreting things in a skeptical hard determinist way.

At the same time as going too far and rejecting plausible compatibilist interpretations, I also argued that Pereboom did not go far enough, and did not sufficiently recognize and acknowledge the injustice involved in practices of punishment. Anything short of funishment would not address the injustice inherent in treating harshly (and much worse than others) people whom—whatever they may have done—HD views as fundamentally morally innocent.

Desert is a way of justifying divergence from a moral baseline. If desert becomes impossible, this does not mean that the moral baseline has somehow disappeared, and "anything goes" (cf. Smilansky 1996). If the moral baseline is that everyone ought to be treated as innocent unless proven guilty (through his or her free actions), then, if no one can be considered guilty, the moral baseline of innocence remains. Not to respect it would be unjust. No one can be guilty, and indeed according to HD everyone is, in the deep moral sense, innocent, and ought to be treated as such, whatever he or she does. That is one of the deepest insights and moral contributions of HD, if indeed we lack free will and moral responsibility. It is thus a mistake to believe that since HD rules out desert, there is no justice (or injustice) in a hard determinist world.

Similarly, if we assume a moral position requiring that a state treat all of its citizens equally unless they come to deserve otherwise through their free actions, then—if it is thought that they cannot come to deserve otherwise—the baseline of equality is all that remains. A state, then, cannot treat some of its citizens substantively worse than it treats others (and, if it does so in some respect, such as by incarcerating them, it needs to offer such compensation as will right the balance).

Yet, as we saw, funishment would lead to the practical refutation of HD. So HD as a deep moral view should declare its bankruptcy; it cannot follow its conclusions, nor form a

basis for moral and social life in the context of punishment. Joining forces with a utilitarian-like consequentialism risks betrayal of the negative but true insights of hard determinism on universal innocence and the concomitant injustice, and involves hard determinism with the grave difficulties of such consequentialism.

6 Conclusion

In *Free Will, Agency, and Meaning in Life*, Derk Pereboom proposes an optimistic model of life that follows on the acceptance of hard determinism, and rejects familiar notions of free will and moral responsibility. In the context of punishment, the topic of my present paper, Pereboom proposes a model of punishment as the incarceration of the potentially dangerous for the purpose of social self-defence, on the analogy of quarantine.

Pereboom responds to my argument that HD must seek to revise the practice of punishment in the direction of funishment, whereby the incarcerated are very generously compensated for the deprivations of incarceration. I claimed that funishment is a practical *reductio* of HD. Pereboom replies, emphasizing the monetary aspects of my claim and rejecting the idea that HD becomes impractical. However, I claim that he misses a key component of my *reductio*: the idea that moving in the direction of funishment will considerably weaken the deterrence of potential criminals, thus leading to an increase in crime, which results in a situation that is much worse, even according to HD. HD, in other words, becomes self-defeating in practice.

Beyond the challenge of funishment, I raised various other difficulties with Pereboom's hard determinist model. It intuitively gives the wrong reply on cases of one-timers, and could not satisfy the intuitions of the relatives and friends of victims of crime on justice. The "universal excuse" available to even the worse criminals, under HD, is likely to harm the motivation of many criminals in a way that is akin to my past discussion of "The Present Danger of the Future Retrospective Excuse." HD would also be much weaker than its rivals in resisting the familiar dangers of the medical model, and the difficulties I have pointed out in the past of the temptations for the systematic lowering of the standards for persecution and conviction, so that many of the innocent (in conventional terms) will end up punished. Hard determinists have emphasized the drawbacks of traditional views that follow the free will paradigm, but these views are often the safeguards of justice, decency, and human rights.

Our conclusions should lead to a re-evaluation of the compatibilist interpretation of moral life, as a richer, more plausible, and safer interpretation than HD. This needs to be combined with a true hard determinist acknowledgment of the deep injustice and tragedy involved in punishment, in light of the absence of LFW. Such a complex view will come closer to doing justice to notions of justice, morality, and decency.

References

Hart, H.L.A. (1970). *Punishment and Responsibility*. Oxford: Clarendon Press.
Lemos, John (unpublished). Moral Concern about Responsibility Denial and the Quarantine of Violent Criminals.

Levy, Neil (2012). Skepticism and sanction, *Law and Philosophy* 31: 477–493.

Morris, Herbert (1976). "Persons and Punishment," *On Guilt and Innocence*. Berkeley and Los Angeles: University of California Press.

Murphy, Jeffrie G. (1973). Criminal Punishment and Psychiatric Fallacies, in Jeffrie G. Murphy, ed. *Punishment and Rehabilitation*. Belmont: Wadsworth.

Pereboom, Derk (2001). *Living without Free Will*. Cambridge: Cambridge University Press.

Pereboom, Derk (2014). *Free Will, Agency and Meaning in Life*. New York: Oxford University Press.

Smilansky, Saul (1990). Utilitarianism and the 'Punishment' of the Innocent: The General Problem, *Analysis* 50: 256–261.

Smilansky, Saul (1996). Responsibility and Desert: Defending the Connection, *Mind* 105: 157–163.

Smilansky, Saul (2000). *Free Will and Illusion*. Oxford: Oxford University Press.

Smilansky, Saul (2001). Free Will: From Nature to Illusion, *Proceedings of the Aristotelian Society* 101: 71–95.

Smilansky, Saul (2005). Free Will and Respect for Persons, *Midwest Studies in Philosophy* 29: 248–261.

Smilansky, Saul (2011). Hard Determinism and Punishment: A Practical *Reductio*, *Law and Philosophy* 30(3): 353–367.

Sommers, Tamler (2012). *Relative Justice*. Princeton: Princeton University Press.

Vargas, Manuel (2005). The Revisionist's Guide to Moral Responsibility, *Philosophical Studies* 125(3): 399–429.

Vargas, Manuel (2013). *Building Better Beings*. Oxford: Oxford University Press.

Waller, Bruce (2011). *Against Moral Responsibility*. Cambridge, Mass.: MIT Press.