



Further Reflections on The Free Will Debate and Basic Desert: A Reply to Nelkin and Pereboom

Michael McKenna¹

Received: 22 May 2019 / Accepted: 29 May 2019 / Published online: 15 June 2019
© Springer Nature B.V. 2019

Abstract

In my “The Free Will Debate and Basic Desert,” I argued that against a familiar claim in the free will debate: that the freedom in dispute between compatibilists and incompatibilists is limited to the type required for an agent to deserve blame for moral wrongdoing, and to deserve it in a sense that is basic. In that earlier paper, I sought a rationale for this claim, offered an explanation of basic desert, and then argued that the free will debate can persist even when divorced from basic desert. Dana Nelkin and Derk Pereboom then argued against my thesis. In this paper, I reply to their thoughtful criticisms.

Keywords Free will · Moral responsibility · Blame · Desert · Dana Nelkin · Derk Pereboom

I am honored to be the focus of Dana Nelkin and Derk Pereboom’s critical scrutiny of my article “The Free Will Debate and Basic Desert.” Both offer characteristically charitable assessments of my arguments followed by impressive criticisms. In what follows, I will offer only preliminary replies, wishing to register that indeed reasonable minds can differ on these matters. I confess to being moved by much of what each has to say. But as the great philosopher Barkley might say—that is, the former NBA star Charles Barkley—I could be wrong, but I doubt it.

1 Reply to Nelkin

Dana Nelkin raises problems for my assessment of two of the candidate accounts of moral responsibility and blameworthiness. Both eschew basic desert-entailing moral responsibility. I contend that each offers alternative justificatory resources while at the same time retaining a close connection to the traditional free will debate. One is

✉ Michael McKenna
msmckenna63@gmail.com

¹ Department of Philosophy, University of Arizona, Tucson, AZ 85721, USA

Lenman's (2006) contractualist proposal. Another is Wallace's (1994) appeal to considerations of fairness. Nelkin disagrees with my assessment of each. She then challenges the way I and others understand the relationship between what a blameworthy person deserves and the normative grounds for blaming one who is blameworthy. I'll comment on each of these three points.

1.1 Nelkin's Assessment of Appeals to Contractualist Resources

Nelkin rightly worries that certain appeals to contractualist considerations might be too revisionary. She cites Scanlon's (2008) and Heironymi's (manuscript) contractualist proposals as unabashedly revisionary, and so inconsistent with my own contention that contractualist resources could be used to justify our actual moral responsibility practices, and not some revisionary variation on them. I'll grant that she is right about Scanlon (of 2008) and Heironymi. But she also includes Lenman's contractualist proposal. About his proposal, she writes:

[W]hile there are structural similarities in the kinds of sanctioning responses licensed by contractualist views such as Lenman's that have parallels to excuse and justification when it comes to moral responsibility practices, it is less clear that such pictures include *blaming* in a sense that is closely related to responses like the reactive attitudes as seems an important part of our practice... The reactive attitudes arguably commit us to ideas of desert, and the case for this is, at the least, more plausible than that they can be licensed by, or are otherwise essentially connected with, a contractualist understanding. (Nelkin, manuscript)

In a footnote (6), Nelkin further comments on Lenman's proposal, whereby contractors would adopt principles "very much like" ours in terms of excuses and exemptions. But Nelkin then expresses skepticism that these principles would be able to account for our actual practices with "all of their current content" in a way that avoids the notion of desert.

I offer here two related points in reply. First, as I understand Lenman, he *is* committed to a desert thesis. It is just that it is not basic. So one way to appreciate what he is up to is that he provides normative resources of a contractualist sort to help justify our persistence in drawing upon the reactive attitudes we have, and so also on thoughts about what agents deserve.¹ For instance, about the reactive attitudes, Lenman writes:

The difference between those Human Beings who are preeminently worthy of love and respect and those Human Beings who are preeminently unworthy of love and respect are too central and important a governing feature of human life and human relationships for the reactive attitudes involved in recognition

¹ As I noted in "Basic Desert, Blame, and Free Will," this would amount to what Rawls (1971) might call post-institutional desert.

of such worthiness and unworthiness to be, either ideally or in fact, open to abandonment. (2006: 19-20)

Turning to the issue of desert, later Lenman considers Scanlon's (1998: 294) reasons for rejecting a desert thesis, which include the consideration that, as Lenman puts it, "when people morally screw up" one might think, "there but for the grace of God go I" (2006: 23). About such a thought, Lenman contends that on the account of autonomy he endorses, and given his contractualist proposal, for Human Beings, "it is plausibly reasonable to think this gamble is a worthwhile and reasonable one for them to make" (23).

To be fair to Nelkin, I take it that her point is that even if this is so, the contractualist anchor used to justify the role of our reactive attitudes is likely not to preserve in full our intuitions and practices that are *directly* about what culpable wrongdoers deserve. Hence, her remark about "all of their current content." Fair enough. But this leads to a second point. What I contend, and what Nelkin rightly interprets me as contending, is that our actual practices can remain our actual practices while at the same time undergoing some minor alterations. This is needed to avoid the charge that any variation amounts to changing the subject in the fashion of a revisionist proposal. If so, these very practices might be plastic enough to remain *those very same practices* even if some of the actual aspects of them could be shed or treated as ancillary or unnecessary. The thought here is simply that with contractualist resources we might be able to find a justification for our actual blaming practices, including some notion of what agents deserve, where this justification requires a strong freedom condition but does not invoke the notion of *basic* desert.

1.2 Nelkin's Assessment of Appeals to Fairness

Now consider Nelkin's remarks about fairness. I claimed that Wallace's (1994) appeal to fairness provides rich resources for justifying our actual blaming practices, and these resources can in turn be used to invoke a robust freedom requirement. Nelkin is correct that one set of considerations regarding fairness are inapt: those involving considerations of fair distributions. Her appeal to intuitions about moral luck seem decisive on this point. (I'll not rehearse her excellent argument here.) However, Nelkin then notes that another consideration of fairness is reasonableness. It is this notion of fairness that Wallace invokes. Here, Nelkin's contention is not that fairness will fall short of what desert might offer as a justification for our blaming practices. She is prepared to consider the possibility that reasonableness could do the needed justificatory work. Instead, her charge is that reasonableness might depend on desert. Nelkin then proceeds to explore a novel thesis about the relationship between reasonableness (in the form of fittingness) and desert, holding that while distinct, they are essentially connected. Momentarily (in Sect. 1.3), I'll comment on her thesis. But here I wish to offer first a commentary on one way to appreciate reasonableness as a feature of fairness, followed by a reflection on what desert is in these contexts regarding blameworthiness. My intention is to try to show that there is a good bit of light between an appeal to reasonableness as it bears on

fairness and an appeal to something like basic desert. (This will come up again in my reply to Pereboom.)

Consider the role of reasonableness in an understanding of fairness. There is a fairly thin way of understanding it, one that helps to cast light on its relation to the distinct distributive considerations that fairness informs. This will include distribution of bearing the burden of sanctions for culpable wrongdoing. Suppose there are rules or procedures falling out of a fair contract that include judgments about what distributions of various benefits and burdens would depend on the behavior of agents who entered into it. It is unreasonable to apply such rules and procedures to agents for whom it is pointless to require compliance—just in the same way that it is unreasonable to expect a toaster to calculate my taxes, or a toddler for that matter. I take it that thoughts about desert go beyond this lean notion of reasonableness, as I will now explain.

So, what makes it so that I can avoid Nelkin's charge that reasonableness depends upon desert? One proposal might be that desert in these contexts is most fundamentally a matter of rightness or justice. It is a matter of doing right by one. It is a matter of what is owed to a person in virtue of how she comports herself. This, it seems to me, involves more than the thin notion of reasonableness identified in the preceding paragraph. Demanding of my toaster that it solve my math problem or that my four-year-old carry the barbells to the attic is unreasonable in a way that need not be thought of as invoking considerations of (moral) rightness or justice. It's just not reasonable, give what we know about toasters and four-year-olds. Or at any rate, I assume it is the burden of one wishing to deny this to show how mere reasonableness does invoke some notion of desert. Yet another more substantive notion, one that I favor, is that desert in contexts involving blame entails that it would be noninstrumentally good that a culpable wrongdoer is rendered worse off or made to suffer in some way as a response to her culpable conduct. Arguably, this way of thinking about basically deserved blame goes beyond what is involved in the notion of reasonableness.

In light of the above, I remain convinced that appeals to consideration of fairness in the form of reasonableness might provide us a way into the free will debate by resources that are independent of basic desert. (Below I will qualify my position further in response to Pereboom.)

1.3 Nelkin on Distinguishing Fittingness from Desert

Now consider Nelkin's striking thesis that the traditional way of framing this debate is misguided. Along with philosophers like Pereboom (2001, 2014), I have assumed that the relevant notion of basic desert links blameworthiness to blame so that blame is *what is deserved*, and with it the attendant harms that can accompany blame. But Nelkin challenges this widely shared assumption. She distinguishes conditions for two different things bearing on moral responsibility: accountability and desert. According to Nelkin, blame, and especially directed blame, bears directly on accountability inasmuch as directed blame involves holding others to account. Agents who satisfy her proposed Quality of Opportunity conditions, she argues,

render holding them to account by blaming them apt, and these conditions *also* satisfy the conditions required for agents to deserve harmful consequences for their bad acts. So far, these claims do not depart from orthodoxy. But then Nelkin proposes the following exciting suggestion: The normative basis for blame is *fittingness*, not desert. The normative basis for harmful consequences of one's culpable wrongdoing is *desert*. Finally, *these can come apart*.

In reply, I wish to exercise a good deal of caution. This is an impressive and creative proposal for rethinking this entire debate. It merits thorough exploration. I have no desire to reject Nelkin's proposal outright. I would also note that to do it full justice requires attention to details I think it best not to explore here. In what follows, I offer a few reasons to regard Nelkin's proposal with caution.

First, for obvious reasons, I sympathize with the idea that there could be a normative basis for linking blameworthiness and blame by means other than desert. At first blush, fittingness is a viable option. Indeed, one might think my own conversational theory of moral responsibility could be enlisted to support Nelkin's contention. On my view, blame as an analogue to a conversational reply invokes norms of conversational intelligibility or meaningfulness, and this is one way to help think of blame as fit for a blameworthy act. So far so good. But as I have also argued in various places (e.g., 2012), mere conversational intelligibility or meaningfulness as a way to unpack the special relation of fittingness seems inadequate for reasons that, apparently, do not seem to be so much of a concern for Nelkin. My worry is precisely that directed blame is liable to harm, and so a norm of intelligibility as a kind of fittingness will not provide proper warrant for an activity that is liable to harm a person, even if the person is a culpable wrongdoer. It is for this reason that I had assumed some sort of moral or ethical relation was needed as a supplement to mere (conversational) fittingness. And it was just the contention of "The Free Will Debate and Basic Desert" (manuscript) that there are other normative resources beyond basic desert to do the normative work needed here and yet do so in a way that invokes the traditional worries about free will.

Second, consider now the thing that, as regards the fittingness of blame, I say, is not so much of a concern for Nelkin and is of paramount concern for me. I worry that directed blame might harm the culpable wrongdoer and so cries out for a further normative basis beyond mere fittingness. Nelkin, by contrast, worries that because directed blaming might *not* harm the wrongdoer, even if fitting, it cannot be the thing that the wrongdoer deserves insofar as what is deserved for culpable wrongdoing is to suffer some sort of harm. Why? The contingency of our blaming practices and the vagaries of the psychology of those blamed make it too much of an iffy affair to claim that being blamed, and so harmed by blaming, is what the culpable wrongdoer deserves. We might blame in especially fitting ways, and yet a guilty party might only profit from our treatment. Indeed, she might get a kick out of it. Others have also recently focused upon the contingency of the link between directed blaming practices and the liability to harm a culpable wrongdoer. Carlsson (2017), for instance, has recently rejected the blame of others as the thing that a blameworthy person most fundamentally deserves for just this reason. Call this *the contingency worry*.

I confess, I do not feel the pull of the contingency worry. Suppose Sofia wins the race and worked the hardest to be the fastest. She clearly deserves the prize for

winning. Grant that. As she walks to the podium to get the award, a piano falls on her head, or she explodes, drops dead, or turns into a turnip. She still deserved the award. The world did not cooperate. That happens. It seems to me that the theorist committed to basically deserved blame can say with no worries about the contingency at issue here that one of the things that a blameworthy person deserves is to be blamed by others in a way that is liable to harm those who are blamed in that manner. If features of a person's character mask or fink the effect of being harmed as a (virtuous) person typically would, well then they failed to get something that they deserved. Not all blaming succeeds.²

Third, I also have some reservations about how Nelkin and others (e.g., Carlsson 2017) conceive of fittingness.³ They seem to think of desert in this context as distinct from fittingness: desert is one thing; fittingness is another. As I think of it, fittingness is a thin notion of rationality that links certain types of responses to evaluative domains, like humor as a response to the funny. Just as humor has its own internal norms, so that an amused response is fitted for a funny joke, I think of desert as a species of fittingness. The "thickness" of an evaluation of a claim of fit is thus supplied by *the internal standards of that domain*. At least as I understand it, merely asserting that an instance of directed blame is fitted for a bit of blameworthy conduct does not inform us of *how* it is so fitted, unlike, say, the way belief is fitted for what is credible, or desire for the desirable. A natural proposal for blame being fitted for blameworthiness is that it is so fitted *because it is what a blameworthy person deserves*.

Fourth, it is also worth noting that one way to avoid the contingency problem and preserve the link between the fittingness of blame and the deservingness of blame is by way of a form of blame that does not make the link between blame and suffering harms contingent. This is Carlsson's strategy (2017). In particular, Carlsson argues, if what a blameworthy agent most fundamentally deserves is self-blame and the pain of guilt, those are essentially harmful. Now, as I have argued above, I am not moved by the contingency worry, but if one were, Carlsson's strategy would offer a way to address this concern.

In closing this reply to Nelkin, I again wish to emphasize that I do not mean simply to reject her fascinating idea here. Decoupling questions of what a blameworthy person deserves from the normative basis for blaming her is an exciting development meriting more attention.

² Note that when discussing the prospects for an analysis of desert, Joel Feinberg wrote, "The deserved object must be something generally regarded with favor or disfavor even if, in some particular case, it is regarded with indifference by a person said to deserve it" (1970: 61). I am grateful to Randy Clarke for calling my attention to this passage.

³ I admit that I might misunderstand those who use these terms, and I also admit that what follows is not something about which I have a high degree of confidence.

2 Reply to Pereboom

Derk Pereboom raises three concerns for my proposal. First, he argues that a basic notion of fairness as it bears on the question of justifying directed blame or praise is indistinguishable from basic desert. If so, then I am wrong to propose that considerations of fairness can give rise to metaphysical worries about freedom in a manner that is substantively different from how they might arise from basic desert. Second, Pereboom reports that on his considered and long-held view, it is an empirical question whether our actual moral responsibility practices, including the practices of praising and blaming, can or are justified on, say, consequentialist or contractualist grounds. Given this, I am wrong to think that what is at issue is whether there are alternative bases for justifying our actual practices in a way that implicates a metaphysically robust freedom condition. Third and finally, as Pereboom sees it, the non-desert based grounds for justifying some sort of moral responsibility practices, the sort clearly not at issue in haggling over the metaphysics of the free will debate, are not lesser or watered-down notions. They are instead profound and of the greatest importance. I'll comment on each of these three points.

2.1 On the Profound Importance of Competitor Notions of Moral Responsibility

I begin with Pereboom's third point. Here I can be brief. I chose my words recklessly. I had no intention of characterizing any possible (metaphysically unproblematic) alternative justificatory basis for our blaming practices as *normatively* lesser or *morally* watered-down. I wish to make no claims at all about their profound importance in this sense. Perhaps a slightly reworded formulation of the offending paragraph would help. Here is what I should have written:

Many contemporary philosophers working on the related topics of free will and moral responsibility contend that the proper way to understand the free will debate is in terms of basic desert-entailing moral responsibility. On their view, attempting to settle the free will problem by reference to any alternative, less metaphysically demanding notion of moral responsibility amounts to nothing more interesting than changing the subject, and hence avoiding the traditional debate altogether. Why? Given these (allegedly) less metaphysically taxing notions of moral responsibility and the weaker metaphysical freedoms they require, there is no reason to think that determinism would be a threat; there's just no philosophical problem to solve. This is how Derk Pereboom approaches the debate (e.g., 2001, xxi–xxii), and many others follow him on this point.

I hope this is enough to clear up any confusion on this point. I suppose it bears mentioning that on this point, I only wish to remain agnostic about the normative or moral matters. Indeed, it is likely that others would protest that these metaphysically unproblematic justifications *are* normatively lesser or morally weaker. Kane (1996), for instance, contends that our lives would not have the value for us that they have if

we were not free in a way that mattered for our being truly responsible for our conduct. So it is perfectly understandable that Pereboom would interpret my remarks in a similar manner. In any event, I have no interest in engaging in this debate, at least not in the present context.

2.2 On the Commitments of Our Actual Blaming Practices

Pereboom disagrees with my focus on the set of practices bearing on the application of our concept of moral responsibility, and in particular our application of the concepts of blameworthiness and blame. I contend that we can focus on these so as to fix upon a substantive issue that can then divide theorists regarding whether those very same practices can be justified by way of resources that require pertinent freedom conditions. Pereboom disputes this:

My position, accordingly, is that it's open that our actual practice of holding responsible, including our practice of justifying penalties, is in turn morally justifiable on the basis of consequentialist or contractualist considerations. I also think that this does not conflict with my incompatibilism.... Thus, I disagree with the claim that the shared domain of practices can play the key role in providing the substantive division between compatibilists and incompatibilists. (Pereboom, manuscript)

Pereboom contends, rightly to my mind, that this is an empirical question that cannot be solved by philosophy alone. Much turns on details of what the actual consequences might be, or instead the rational options for a party to a contract.

In reply, perhaps it will be useful to begin by rehearsing what led me to my proposal. Consider Pereboom's long-standing thesis (e.g., 2001) that various candidate proposals for justifying our blaming practices do not bear on the metaphysical issues at stake in the free will debate because (1) they do not rely upon a notion of basic desert, and (2) the justificatory resources they offer are so transparently compatible with determinism that there just is no significant basis for philosophical perplexity. The charge then proceeds that these proposals cannot be deployed in the service of solving the free will problem, since they would only do so by changing the subject. In my current efforts to resist this bit of orthodoxy (manuscript), I then propose that we consider justifications that accept (1) but deny (2). To avoid the charge of changing the subject, I must identify something other than basic desert that can serve as the shared domain of dispute between, say, compatibilists and incompatibilists. Otherwise, as Pereboom explained (manuscript), we are just involved in a verbal dispute. It is in this context that I contend that we can plausibly treat the shared domain of dispute between compatibilists and incompatibilists in a way that is not limited just to views that rely upon basic desert; we look to our actual practices. Pereboom (manuscript) then rejects my proposal by noting that his own free will skepticism is consistent with a justification of our actual blaming practice. So, he reasons, these very practices cannot be determinative of the substantive issues that divide compatibilists and incompatibilists regarding the metaphysics of the free will debate.

Here, in reply, I only wish to make two points. First, Pereboom is clearly correct about this point: It is conceivable that there could be an adequate justification of our actual blaming practices (and not just some revised version of them) that is compatible with his free will skeptical conclusion. Whether or not this is so is an open empirical question that is not within the domain of philosophy to settle. I never meant to deny this. Indeed, doing so would be reckless, since there are perspicuous contenders in this arena. Pereboom mentions Dennett's proposal (1984: 66). But there are other familiar ideas, such as one often offered in criticizing Strawson's (1962) practical rationality argument (e.g., see Wolf 1981). Strawson contends that if we could rationally choose whether to retain our reactive attitudes and attendant practices, we could only do so in terms of the gains and losses to human life. But, so the objection goes (Wolf 1981), this would only establish that we would do best to treat people *as if* they are morally responsible and free in the sense implicated by our commitment to the reactive attitudes, even if as it happens they are not.

Second, granting the preceding point is consistent with my proposal. Note that the candidate justifications that Pereboom contends might be compatible with his free will skeptical conclusion are proposals that would commit one to an illusionist thesis, or the need for an error theory of some sort. This is how Pereboom describes Dennett's (1984: 66) view, which Pereboom offers as a counterexample to my proposal. He writes:

Even if the claim that we are morally responsible cannot be justified, there may be a practical argument for nevertheless treating ourselves and others as if it were true. (Pereboom, manuscript)

All that is needed, then, to preserve my proposal is that when attending to our actual blaming practices, we fix upon justifications that treat a proper subset of the judgments expressed or implicit in our practices as reliable, true, or well-grounded. Given this background assumption, we can then evaluate alternative justifications that employ robust freedom conditions.

Of course, to avoid the charge of mere verbal disagreement, there will still be the matter of seeking justifications of our actual blaming practices that under careful scrutiny do not rely upon the assumption of basic desert. And as is suggested in my quotation of Pereboom, which I will repeat here, that also creates problems:

The free will skeptic will call into question any blaming practices that presuppose that the agent being blamed is morally responsible in the basic desert sense. Since much actual human behavior has this presupposition, any skeptical account of blame will be revisionary. (Pereboom 2013: 190)

By the same reasoning, by Pereboom's lights, any non-skeptical account of blame that did not deploy the concept of basic desert *would* be revisionary. However, I have countered that we can theorize about our actual—*rather than a significantly*

revised set of practices—by treating the basic desert presuppositions in our actual practices as not necessary or essential to them.⁴

2.3 On the Prospects for Theorizing in Terms of Basic Fairness

Now consider Pereboom's intriguing remarks about fairness. I proposed that appeal to considerations of fairness provide an alternative to basic desert wherein the justification of our actual blaming practices in terms of fairness still helps us make sense of the free will debate. I then cited Wallace (1994), who has argued in these terms. Pereboom in reply claims that my appeal to Wallace is misplaced because Wallace's remarks were about distancing his appeal to fairness from retribution, not basic desert. Pereboom then proceeds to identify the key notion of fairness that he contends could play the pertinent justificatory role, basic fairness. He then argues that this notion is not different in any relevant respect from basic desert. Hence, my claim about appeals to fairness being a genuine alternative path to the free will debate fails; it's the same path. To proceed, I'll consider each of these separately: Pereboom's assessment of Wallace's appeal to fairness, and then Pereboom's positive proposal.

Regarding my appeal to Wallace, it is true that at various junctures in his (1994) Wallace expresses his position as one that is an alternative to retribution rather than basic desert, but the rationale he offers to distinguish his view from retribution applies more generally to basic desert. For instance, Wallace claims that we can blame without believing that it is intrinsically good that "a person should suffer some harm" (Wallace 1994: 60). Consider also this passage:

It is true that I have not provided a basis for the strong conclusion that wrongdoers positively deserve to suffer the harms of moral sanction in this way. (Wallace 1994: 227)

This quotation is expressed in terms of desert, not retribution, and so *seems* to be decisive in my favor as regards Wallace's considered view. However, matters are more complicated, and some of what Wallace writes seems rather to support Pereboom. Most crucially, consider how Wallace expands upon the previous remark just quoted:

I have repeatedly urged against building such a retributivist interpretation into the very stance of holding people responsible... Furthermore, the principles of fairness I have articulated provide no support for these retributivist conclusions. Those principles are negative in form, identifying basic necessary conditions for the fairness of holding people responsible. They say that it is not reasonable to hold someone accountable who lacks the power of reflective self-control, and that people do not deserve the responses of blame and moral

⁴ Wallace also suggests that a certain interpretation of our practice (a retributive one) need not be thought of as "...necessarily embedded in the self-understanding of ordinary participants in the practice" (1994: 60).

sanction if they have not done anything wrong. *It follows from these principles that wrongdoers who possess the powers of reflective self-control deserve the response of blame and moral sanction, in the sense that it would not be unfair to respond to them in these ways.* We may say that no moral complaint can be lodged against such treatment, on grounds of fairness, if the conditions of accountability and blameworthiness I have identified are fulfilled. But this is weaker than the conclusion that we have a positive moral obligation to inflict suffering on wrongdoers, or that such suffering would be morally good. (1994, *my italics*: 227)

Reflecting on Wallace's comments, it seems that both Pereboom and I have reasons to hold that Wallace would side with each of us. Of course, my concern here is not with scholarship, but with what we might learn from Wallace's thinking about fairness and desert.

In Pereboom's favor, as is featured in the italicized sentence in the quotation above, Wallace explicitly state that wrongdoers who possess pertinent powers deserve responses of blame and moral sanction. But in my defense, note that Wallace qualifies the desert thesis he endorses so that it just amounts to the contention that "it would not be unfair to respond to them in this way" (277). And, moreover, he states that he does not offer a positive justification for concluding that wrongdoers deserve to suffer. As I have argued, I understand appeal to basic desert to be precisely designed to offer a positive justification for blaming and for the harms that attend blame. Indeed, this is certainly how Pereboom has understood the notion as well.

Do I therefore get to claim that my reading of Wallace's is the correct one, and that his conception of fairness is distinct from basic desert? In fairness to Pereboom, I do not think the passage above unequivocally supports this conclusion. This is because Wallace's remarks also link the notion of a positive justification to *positive obligations* to issue sanctions. But it does not seem to me plausible to contend that basically deserved blame of the sort at issue in this debate involves a positive duty or obligation to blame and harm culpable wrongdoers. In this sense, the notion of a positive desert thesis as Wallace is thinking of it is stronger than one that I would take to be defensible. Nevertheless, a slightly weakened, Wallace-inspired, notion of a positive desert thesis of the sort I favor would *still* count as stronger than the mere negative one that Wallace identifies from considerations of fairness. This positive thesis renders it merely permissible for a blamer to blame a wrongdoer, but it appeals to the noninstrumental goodness of blaming and so harming the wrongdoer as a positive justification for the blaming activity.

Working with the preceding Wallace-inspired notion of a positive desert thesis for blame, we can now get a bead on a fairness thesis for blame as Wallace understands it. It is one that does *not* commit to there being a positive justification for blaming, nor to the noninstrumental goodness of any harms involved in doing so. While it allows inferences that a culpable agent might deserve blame, it is just in the negative sense that it would not be unfair to blame a person. Given this notion, I remain committed to the contention that this notion of fairness has the resources to

give rise to credible metaphysical worries about free will of the sort traditionally in dispute between compatibilists and incompatibilists.

Now consider Pereboom's own positive proposal for a notion of basic fairness. He contends that it has all the features that basic desert has, and so does not offer a substantive point of difference between us. The first point I wish to make here is that if Pereboom is correct that there is a conception of basic fairness that is at least extensionally equivalent in all of its judgments to those generated by way of basic desert, that does not show that the weaker notion of fairness I have identified above, the one I attribute to Wallace, can be given the same diagnosis. Hence, I contend, there is a *bona fide* notion of fairness which is sufficiently distinct from basic desert. This is so even if the conception of basic fairness to which Pereboom appeals is not sufficiently distinct from basic desert. (Recall that by way of Wallace's notion of fairness, in Sect. 1.1 I argued that we could appeal to a minimal notion of the reasonability of holding agents to expectations, one that is distinct from the notion of reasonableness that might be at work in considerations of basic desert. This figured in my reply to Nelkin.)

This first point in reply to Pereboom's positive proposal is not altogether satisfactory. One might object that there is a substantive dispute here about what fairness actually is. Even if the notion Wallace appeals to differs from the one Pereboom identifies as basic, the key question is whether Pereboom's rather than Wallace's gets to the heart of what fairness is. After all, Pereboom might just point out that Wallace failed to make an important inference to which he was entitled. Fairness, when conceived as basic, *does* warrant a positive justification for delivering the harms of blaming, and it *does* justify its being noninstrumentally good. How should I respond?

Perhaps readers expect me to resist. If so, I fear I will disappoint. Pereboom's proposal strikes me as promising and deeply insightful. He may be correct. I do not wish to unilaterally rule this out. If there is a notion of basic fairness that operates as Pereboom contends, then I want to grant him the point. Of course, and this leads to my second point, I would add that in the broader context of the debate between us, I also identified metaphysically interesting freedom presuppositions in Lenman's (2006) contractualist proposal. I would say the same about Scanlon's (1998) contractualist appeal to justify substantive responsibility. So even if Pereboom is correct about the issue of basic fairness, my wider thesis still has some life in it. Thus, in what follows, I will just raise two tentative reservations about Pereboom's notion of basic fairness and simply leave the dispute between us unresolved.

First, as a conceptual point, unlike desert, fairness seems to be most fundamentally a comparative notion that fixes upon the treatment provided to a recipient who can then be said to be treated fairly or unfairly in relation to how others are or would be treated. *Desert is not comparative in that way*. It is settled by some desert-base that applies to the deserved person and then a response that is fitted for it. Of course, it is natural to think that it would be unfair for another who also deserved a similar benefit or harm not to likewise get the same thing that she

deserves. But that is a further thought about fairness that is not conceptually fundamental to desert, or so it seems to me.⁵

Second, there is one sort of case at the limit that does seem to yield different judgments. If so, it meets Pereboom's challenge to me to show that there could be impositions justified by one notion but not the other notion. Pereboom considers the possibility that God creates people with the same talents and material goods and that this can be seen as fair albeit not deserved. Granting that in this sense there *is* a difference between fairness and desert, Pereboom then argued that basic fairness and basic desert are essentially the same, since the notion of basic fairness is not about the fairness of distributions, as it would be in the case of God's equal distribution of talents and material goods. But now, consider another sort of case, not with several individuals but just one. God creates a world with only one agent, Zappa, imbued with talents and abilities and motives to do various things, some of which would please God and some that would defy him. Suppose Zappa does various things, some of them pleasing to God, some of them offensive. They include great works of art and, say, defiling some of God's natural beauty. It is plausible to think that if God were to reward and to punish Zappa, to praise him and blame him, it would be deserved. But would it be fair to Zappa? Perhaps intuitions differ here, but it seems to me that without a comparative class of agents who might or might not be treated likewise for similar conduct, judgments of fairness are misplaced.

Are the two preceding points decisive? No. Maybe Pereboom is right about this. I leave it to the reader to consider the matter further.

3 Closing Considerations: Why Does this Matter?

In closing, it is worth asking why bother with this debate. Why make such heavy weather about whether the classical metaphysical issues regarding the free will debate hang exclusively on the notion of basic desert? Who cares? Perhaps to outsiders this debate is a perfect example of an angels-dancing-on-the-head-of-a-pin exchange. Here is why I think it is not. The free will debate has a significance in our understanding that is resistant to many efforts to dispense with it. There's a reason that it is one of the perennial problems of philosophy. Given this, there is a risk of linking it exclusively to the presupposition of basic desert, since that notion might be problematic for reasons that have nothing to do with skepticism about free will. Suppose, for instance, that basic desert in the context of blame

⁵ Perhaps this is too quick. Surely some desert judgments *do* invoke comparative judgments. The fastest person deserves to win the race, and this requires a comparison to the others racing. While there is some force to this objection, it can be accommodated. In these cases, the comparison to others is only one defeasible factor. Notice, for instance, that the first person to cross the finish line in a fair race might *not* deserve to win. Another person, who might have been faster and trained harder, might have suffered an unexpected bout of illness or been stung by a bee. Still, it would be fair to assign the win to the first person to cross the finish line. (I am indebted to Robert Wallace for raising this issue.)

and punishment does entail that it is intrinsically or noninstrumentally good that culpable wrongdoers suffer. Now also suppose that this axiological thesis is philosophically indefensible. Then the conceptual resources required to make sense of the free will debate are defective from the get-go. Entertaining alternative normative bases for justifying our blaming practices that also help us make sense of the metaphysics of the free will issue guards against this outcome.

Acknowledgements For helpful comments, I would like to thank Dana Nelkin, Derk Pereboom, Carolina Sartorio, and Manuel Vargas.

References

- Brink, David O. and Dana Kay Nelkin (in preparation). The Nature and Significance of Blame. In *Oxford Handbook of Moral Psychology*, ed. John Doris and Manuel Vargas. Oxford: Oxford University Press.
- Carlsson, Andreas Brekke. 2017. Blameworthiness as Deserved Guilt. *Journal of Ethics* 21: 89–115.
- Dennett, Daniel. 1984. *Elbow Room: Varieties of Free Will Worth Wanting*. Cambridge: MIT Press.
- Feinberg, Joel. 1970. *Doing and Deserving*. Princeton: Princeton University Press.
- Heironymi, Pamela. Manuscript. *Minds that Matter*.
- Kane, Robert. 1996. *The Significance of Free Will*. Oxford: Oxford University Press.
- Lenman, James. 2006. Compatibilism and Contractualism: The Possibility of Moral Responsibility. *Ethics* 117: 7–31.
- McKenna, Michael. 2012. *Conversation and Responsibility*. New York: Oxford University Press.
- Nelkin, Dana. Manuscript. Desert, Free Will, and Our Moral Responsibility Practices.
- Pereboom, Derk. Manuscript. What Makes the Free Will Debate Substantive?.
- Pereboom, Derk. 2014. *Free Will, Agency, and Meaning in Life*. New York: Oxford University Press.
- Pereboom, Derk. 2013. Free Will Skepticism, Blame, and Obligation. *Tognazzini and Coates*. Eds. 2013: 189–206.
- Pereboom, Derk. 2001. *Living Without Free Will*. Cambridge: Cambridge University Press.
- Rawls, John. 1971. *A Theory of Justice*. Cambridge: Harvard University Press.
- Scanlon, T.M. 2008. *Moral Dimensions: Permissibility, Meaning, Blame*. Cambridge: Belknap Harvard Press.
- Scanlon, T.M. 1998. *What We Owe to Each Other*. Cambridge: Harvard University Press.
- Strawson, P.F. 1962. Freedom and Resentment. *Proceedings of the British Academy* 48: 187–211.
- Wallace, R.Jay. 1994. *Responsibility and the Moral Sentiments*. Cambridge: Harvard University Press.
- Wolf, Susan. 1981. The Importance of Free Will. *Mind* 90: 386–405. (Reprinted in Fischer and Ravizza, eds. 1993).

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.