Responsibility, Methodology and Desert

Manuel Vargas

Introduction

According to one recent approach to moral responsibility, the justification of our responsibility practices—roughly, practices of fault-finding and the ensuing license to justifiably feel and oftentimes express moralized blame—is in broadly instrumental terms (Vargas, 2013). That is, the justification of responsibility is to be found in the way it helps build us into better beings of a particular sort. The approach has also been developed in tandem with a distinctive methodological framework, according to which the account of moral responsibility is to be understood as a revisionist one, requiring principled departures from some putative aspects of commonsense thinking about moral responsibility. So, there are two pieces of the theory: a story about the normative foundations of moral responsibility, and a story about the methodological picture that animates that story. These pieces are separable. One might accept an instrumentalist theory of responsibility without being a revisionist, and one might be a revisionist without being an instrumentalist.

This article focuses on two issues that have been raised against this account. The first concerns the question of whether we do better to put aside revisionist theorizing in favor of a putatively distinctive *ameliorative* analysis. The second concerns whether a "moral influence" or instrumentalist approach to responsibility can accommodate the central issue in debates about moral responsibility, namely the *basic desert* sense of responsibility. The present article

argues that, first, the proposed ameliorative analysis is not properly understood as a competitor to revisionist theorizing, and second, basic desert can be readily accommodated within the instrumentalist approach under consideration.

The first section canvasses the normative structure of the aforementioned theory of moral responsibility. The second section unpacks the methodological framework that informs that account. In the sections that follow, the afore-mentioned issues—ameliorative analyses and basic desert—are taken up and addressed.

Sketching a Theory

Consider the following account of moral responsibility. Whether someone is morally responsible for some bit of behavior is a function of two things: first, whether they are the right kind of agent—a responsible agent—and second, whether in so behaving they have violated justified norms of moralized praise and blame. What makes someone a responsible agent? The person must be suitably responsive to reasons, and in particular, the moral reasons that were relevant in that context of action. This is a power frequently lacked by the very young, and those afflicted with cognitive and affective impairments of various stripes.

Responsible agency is a variable feature of agents. In some contexts, and with respect to some reasons, a person might not be a responsible agent, and thus, might not be a proper target of moral blame. In other contexts, or with respect to other reasons, the same agent might well be the right kind of agent, and thus, a candidate for deserving of blame. In this way, the young, the old, and most everyone in between may find themselves on either side of the line that divides responsible from non-responsible agency.

We have a story about the kinds of agents we rightly blame, but what norms are there about blame among such agents? The norms of praise and blame have a discernible structure. The norms of *moral* blame, at least, track whether or not agents have demonstrated due concern for what moral reasons there are (or alternatively, what moral reasons we collectively take there to be).

This network of practices, attitudes, and norms characteristic of moral responsibility is something we can call *the responsibility system*. Our task is not to chart the responsibility system's general contours in the manner of an anthro-pologist seeking to describe a set of social practices. That may have some utility, but our principal interest is to explain whether and why rules of that sort have any claim on us when they do. There are lots of social practices with attendant rules—games, for example—that we do not take to have any claim on us, unless we voluntary submit to those rules. For the responsibility system to have the kind of authority it presents itself as having in our lives, we would need some explanation of the nature and source of its normative significance.

The story of the responsibility system's justification, and its corresponding authority in our lives, derives from two sources. First, it has justification *as a practice* to the extent to which it produces the right kinds of effects without unacceptable cost. The responsibility system, with its attendant entrenchment of social and moral norms, and its practices of praising and blaming, is justified if having these practices helps develop, sustain, and extend our ability to rec-ognize and suitably respond to moral considerations. It is plausible that our practices frequently satisfy this demand.

Importantly, the instrumental justification for having a responsibility sys-tem need not, and plausibly cannot, exclusively produce instrumental norms at the first order. Foul calls in sports provide an instructive model. The justifi-cation of a system of foul calls is typically instrumental. Having fouls in a sport is justified because of the way it helps preserve the safety of the players, and because it does so in a way that tends to minimize the negative effects on the enjoyment of the fans and the participants of the sport. It does not follow from this that it is an instrumental matter whether a foul should be called in this case or that. Rather, whether a foul is deserved is a matter of what the rules are, and what the rules say about the particular case at hand. Foul calls that only tracked whether so calling the foul would produce desirable effects would, by and large, be a mess. Unless, of course, it was pursued in the rule-regimented way described, which is to say, as practice that is not itself instrumental in structure and internal functioning, but for which there is an instrumental justification for it, external to the practice.

There may well be contexts where differences in otherwise tacit back-ground conditions undermine the ordinary justification for having the practice. Under conditions of widespread social distrust, fundamental moral disagree-ment, or in encounters with agents who have very different psychologies than ours, we should expect that this justificatory warrant is particularly prone to being defeated. Again, though, the model of foul calls is instructive. Foul calls presume a set of background conditions. If there is a lightning storm, or the enforcement of the rules otherwise jeopardizes the players, the rules lose their customary force. No one, however, thinks that this everywhere vitiates the rules of the sport, or shows that the rules are not in good standing when the back-ground conditions are satisfied. The same is true of moral responsibility.

A second piece of the justification of the responsibility system concerns the interest of individuals in *being* responsible agents. Here, the power of the practice is contingent, but nevertheless deeply rooted in our need for social life, social esteem, and the conditions of cooperative living. Adherence to the responsibility system, being competent at navigating its demands, and having a genuine concern for one's status within the system, is a central part of our having full moral standing in our communities. When one is not regarded as competent at the responsibility system—for example, when one lacks a deep and thoroughgoing awareness of the system's intricacies, or when one lacks a

disposition to respond to it appropriately—one cannot be viewed as competent and reliable partners in schemes of social cooperation and coordination. So, people have both a first-personal and a second-personal interest in agents being competent at navigating the responsibility system. Satisfying these interests tends to require robust internalization of the norms of moral responsibility.

The wider details of this approach, understood in terms of a unified and systematic account of moral responsibility has been pursued elsewhere (e.g., Vargas, 2011a; Vargas, 2013; Vargas, 2015a; Vargas, 2015b). However, it is worth noting that a good deal of its machinery overlaps with other accounts of responsibility. For example, the idea that what characterizes some agents and not others as suitable targets of responsibility practices is reasons-responsive agency, has been defended by a number of prominent accounts (e.g., Brink and Nelkin, 2013; Fischer and Ravizza, 1998; McKenna, 2013; Nelkin, 2011; Wallace, 1994; Wolf, 1990). In its construal of the general shape of norms of blameworthiness, the account resembles features of P.F. Strawson's (1962) "quality of will" account, which characterizes our blaming reactions as responses to perceived failures of due moral concern. In its justification of the responsibility system as a system of moral influence, the above account draws from both traditional consequentialist accounts of moral responsibility (e.g., Nowell-Smith, 1948; Smart, 1961), but also newer accounts that have added nuance to the approach (e.g., Arneson, 2003; McGeer, 2014; McGeer, forthcoming; Miller, 2014a; Fricker 2016). The idea that a practice or institution can have non-instrumental rules that are justified instrumentally is an idea found in Rawls (1955) and Hart (1959). In recent work by Joseph Raz (2011), and in a different way, in work by Christopher Bennett (2002), one finds something like the above explanation of an agent's interest in being held responsible, and in being regarded as a responsible agent. To be sure, few philosophers would accept all the pieces of the proposed account, but the account is continuous with a variety of standard approaches to responsibility.

Some Methodological Matters

One objection to the foregoing picture is that it fails to capture important strands of ordinary thinking about responsibility. For example, there is a prominent philosophical tradition according to which the moral responsibility of individuals requires that they have a kind of free will incompatible with the truth of determinism. In contrast, for all that has been said in the above account, the kind of responsibility offered above seems entirely compatible with the thesis of determinism being true.

Here, two methodological commitments help clarify the stakes. First, what the above account offers, first and foremost, is an account of the normative foundations of our everyday practices of praising and blaming. It is not a central

commitment of this approach that it vindicates how we in fact think about the conditions on moral responsibility. In this, it is akin to normative theories of ethics that can sometimes notably depart from folk understandings of ethics. This approach takes its inspiration from physics, mathematics, and so on, where our all-things-considered, rationally best accounts of the phenomena under investigation can depart from folk convictions about those phenomena.

This does not mean that the present account of moral responsibility is untethered from the world. On the contrary, this approach maintains that it has the best and perhaps the only substantial tether to the world. On this *practice-based* (Vargas, 2004) or *phenomenalist* (Vargas, forthcoming) approach to moral responsibility, the animating methodological conviction is that the stakes of a theory of moral responsibility are primarily our responsibility—characteristic practices and attitudes (blaming, indignation, and the like). Whether a theory that satisfactorily accounts for those things also captures the kinds of theoretically naive, pre-philosophical convictions we might have about our powers is a further, and (at best) secondary matter. Our ordinary convictions can be sacrificed on the altar of a methodology that puts features of real-world practices above the potential phantasms we may posit behind those practices.

The practice-based or phenomenalist approach draws its inspiration from Strawson's (1962) methodology, and from his injunction to avoid the "panicky metaphysics" he found in the claims of incompatibilists. In contrast, libertarians, incompatibilists, and at least a few compatibilists have pursued a *conceptualist* strategy, whereby the aspiration is to identify (typically, from the armchair) and describe our beliefs, concepts, and or ideas that figure in responsibility.

In addition to the normative concern and the phenomenalist methodology, a second methodological feature is notable in the present approach. Building on some of the insights that arise from a normatively-focused, phenomenalist methodology, the present account maintains that any satisfactory account of responsibility is very likely to be *revisionist*. A revisionist theory holds that truths about some x are in conflict with commonsense views about x (cf. Vargas, forthcoming; McCormick, 2016). Thus, revisionism about moral responsibility is the view that truths about moral responsibility are in conflict with commonsense views about moral responsibility.

Revisionist accounts have been offered for a variety of phenomena—propositional attitudes, race, gender, morality itself, etc.—and they remain serious contenders in a variety of philosophical domains. The shared motivation for these views is that the thought that our epistemically best accounts of some domain will need to jettison some of the flotsam of ordinary thinking, flotsam partly constituted by cognitive accretions irrelevant to the nature of the thing under discussion. Detailing the conceptual and cultural encrustations on the normative core of our practices is an interesting project, but a distraction from

whether there is a normatively appealing version of these practices available to us.

The responsibility revisionist's broad thought is this: at least some of the time, a lot of people are inclined to think that moral responsibility for wrong-doing requires powers of agency that we are unlikely to have (or, at least, that we have no positive reason for supposing we have), and that those powers—collectively, we can dub them *metaphysically robust free will*—require things like the ability to do otherwise (given the past and the actual laws of nature) or the ability to be self-caused or unmoved movers, or otherwise ultimate sources of actions. In contrast, the revisionist account requires something more pedestrian: that we are able to recognize and respond to moral considerations, and that that our participating in a system of moralized praise and blame tends to, over time, get creatures like us to do a better job of recognizing and responding to what moral reasons there are. That's the theory. Some objections follow.

Revision vs. Amelioration

In "Two Ways of Socializing Moral Responsibility," Jules Holroyd offers a rich and nuanced discussion of the alternatives available to those who accept the conjunction of two claims. The first claim is *the social thesis*, or the idea that "responsibility is constituted by our social practices, rather than any meta-physically deep notion antecedent to our social relations" (Holroyd, 2018, p. 137). The second claim is the *justification thesis*, or the idea that we [rightly] want to ask "whether and why we might want [those social] practices" (137).

Within the set of theories that accept both the social thesis and the justifi-cation thesis, Holroyd identifies a couple of competitor approaches, modeled on the accounts of McGeer (2015) and Vargas (2013), respectively. Holroyd goes on to suggest that the former approach does a better job of handling cases of moral ignorance than the latter, and partly on these grounds, goes on to offer a framework for developing an account that builds on and extends McGeer's scaffolded responsiveness approach to responsible agency. An intriguing sug-gestion in her development of that framework is the following idea: although both McGeer and Vargas pursue revisionist approaches to the theory of moral responsibility, Holroyd thinks we do better to instead offer an *ameliorative analysis* of responsibility.[1]

Drawing from work by Sally Haslanger (2012), Holroyd characterizes mat-ters this way: "an ameliorative analysis starts by asking what we want the con-cept of responsibility for and what concept will serve those purposes, with no assumption that the answers we give will yield an analysis that closely tracks our existing understanding of moral responsibility" (Holroyd, 2018, p. 138). Put this way, an ameliorative analysis seems entirely compatible with revision-ism—neither approach requires that the result of theorizing be something that closely tracks our understanding of moral responsibility, any more than the chemical theory of water closely tracked folk understandings of water as, say, one of the four basic indivisible substances. In light of this, one might think that ameliorative analyses are simply a species of revisionist theorizing.

Holroyd does not seem to think of things in this way. She goes on to argue that there is an important difference here, because "Revisionist analyses seek to revise our existing concept of responsibility, and so remain anchored in our extant concept" (138). In contrast, "Ameliorative analyses do not try to unpack and articulate our concept. Instead, this sort of inquiry is normatively moti-vated: we start by asking what the legitimate purposes are for which we want and use the concept, and then, having articulated those purposes, we identify which concept we ought to use given those purposes" (157).

For the moment, put aside the question of how we are to establish the legitimate purposes of a concept without appealing to any existing concept. The normative motivation identified by Holroyd is not a unique feature of ame-liorative analyses. As she noted at the outset, a central preoccupation of current revisionist theorizing, at least about responsibility, is that it is centrally con-cerned with the normative foundations of our practices, of explaining whether and why we ought to have practices of that or some other sort. It is, after all, why Holroyd is right to observe at the outset of her discussion that there is agreement about the *justification thesis*, the idea that we must ask whether and why we might want these or those practices. The idea that an ameliorative analysis gets us something more than standard revisionist approaches because it "leads us to explicitly reflect on the question of what we want our concept of responsibility for and what work the concept *ought* to be doing for us" (168) is a putative difference difficult to distinguish from the revisionist injunction to pull apart descriptive accounts of responsibility and the more central and important prescriptive question of what we ought to do, of how we ought to organize our thought and practices (Hurley, 2000; McCormick, 2016; McGeer, 2015; Singer, 2002; Vargas, 2004; Vargas, 2013).

One place to look for a difference is in Holroyd's explanation of her rejec-tion of revisionism. She characterizes revisionism as "essentially conceptual analysis, with a commitment to rationalizing that concept as far as possible" (157). Perhaps there is room for disagreement about whether or not the revi-sionist methodology is essentially conceptual analysis. However, revisionists have tended to understand their accounts as *competitors to*, as opposed to a *species of* conceptual analysis approaches to moral responsibility (Nichols, 2006; Nichols, 2015; Vargas, 2011b). By revisionist lights, it seems a mistake, or at least misleading, to run together a revisionist project—one that is explicitly formulated around *a need to change the concept*—with traditional understand-ings of conceptual analysis, which tends to regard coherence with the existing concept as a hard constraint on an adequate theory of that thing. So, what can be said about these issues?

First, it is important to recognize that the idea that a revisionist theory's concern to begin from where we find ourselves does *not* commit revisionist theories to a rationalization of the concept, come what may. As several revisionists have been at pains to emphasize, we may discover that empirical, conceptual, or normative pressures might well force us to altogether abandon responsibility, and instead, to go in for a replacement notion that is, strictly speaking, *not* responsibility (McCormick, 2016; Nichols, 2015; Vargas, 2013; Vargas, forthcoming). In Vargas (2013), this is the idea of denotational revisionism, a position that McCormick (2016) has defended as a form of "replacementism." And, Nichols' (2015; 2017) *discretionism* holds that the proper reference for "free will" and "moral responsibility" is permissive, allowing us to say truly that someone is *not* responsible while also maintaining that people are responsible. So again, the putative difference between Holroyd's preferred methodological approach and revisionism does not seem rightly located in a difference about whether one can, at the end of inquiry, reject ordinary concepts or practices.

Second, Holroyd may be interpreting revisionism in a broadly *conceptualist* way. This would make sense of her reading of revisionism as robustly tethered to the ordinary concept, but it would do so on pain of rendering revisionism inconsistent with both the methodological origins and main implementations of current revisionist approaches (which tend to be practiced-based or phenomenalist in their methodological orientation). Recall that *phenomenalist* approaches are first and foremost concerned with identifying whether there are adequate grounds for preserving some set of practices, even on pain of departing from our pre-philosophical convictions about (or if you like, *the concept of*) those things. If this is how we understand revisionist approaches—and, for what it is worth, this is how several revisionist accounts understand themselves—then the rejection of revisionism on the grounds that it is essentially conceptual analysis misses some of the distinctive (and Holroyd-friendly) methodological commitments of the approach.

Third, to the extent to which there is a difference in something like the degree of philosophical interest in characterizing and starting with the practices as we find them, this sort of difference seems slender, and if notable, to favor the revisionist in its import. For the revisionist, while it is true that theorizing starts with the everyday phenomena of our practices, and even a diagnosis of current conceptual errors, this is because *it is the rational and moral appeal of these things that just is the subject of the inquiry.* The final account could, and does, oftentimes end up at some remove from where it started. Here's the kicker, though: absent starting with the everyday stuff of responsibility—its characteristic practices, attitudes, and beliefs—it is not clear how it even makes sense to say that the ensuing account *is a theory of moral responsibility.* Ameliorative analyses are all good and well, but if one intends to offer an ameliorative analysis of moral responsibility, there has to be something in virtue of which it counts as a theory of moral responsibility.

In connection with these issues, Holroyd gestures at McCormick's (2013) discussion of two challenges for any revisionist account. The first is what McCormick calls the *reference anchoring problem*, that is, the challenge of explaining why, once we go in for revision, it still counts as a theory of the thing that is supposed to be revised (the reference-anchoring problem). The second is the *normativity anchoring problem*, or the idea that revisionist accounts have to explain where the normative appeal of the account comes from, once it forgoes reliance on the folk concept.

McCormick has identified important challenges for any revisionist theory. However, the force of these challenges largely derives from the fact that the revisionist is inclined to defect from ordinary thinking in the interest of what is normatively appealing. This is why questions about reference go live for the revisionist: in what sense is the revisionist proposal still an account of the concept, and if it involves a shift away from it, why should we think that the proposal has normative appeal? Put differently, on McCormick's treatment of the challenges, these challenges arise precisely because revisionism has the features that Holroyd attributes to her ameliorative approach. So, we might turn these same challenges on the ameliorative analysis: in virtue of what is the analysis at all an analysis of *moral responsibility*, and how does it get normative appeal in a way that is any different than the instrumentalism of many revisionist accounts?

Here, it may be worth noting that McCormick is inclined to think that current revisionist accounts have a pretty good answer to the reference-anchoring problem (McCormick, 2015; McCormick, 2016; McCormick, 2017) and a promising if incomplete line of reply to the normativity-anchoring problem (McCormick, forthcoming). McCormick is more sanguine about revisionism as a kind of replacementism, instead of a reference-preserving story that only purges connotational content, and perhaps trims the untoward features of existing practices. One might reasonably disagree about the prospects of denotational revisionism (Vargas, forthcoming). However, this is mostly an in-house debate among revisionists, rather than a disagreement about whether revisionist accounts have a promising route forward.

Where do things stand? Holroyd says we start an ameliorative analysis by "asking what the legitimate purposes are for which we want and use the concept" (157). So, perhaps ameliorative theories are best understood as revisionist theories that skip—or perhaps just tacitly invoke—what revisionists tend to regard as a preliminary step of trying to provide some account of a received concept for the phenomena in question. It is unclear whether doing without that intermediate step is methodologically appealing. How does one do what Holroyd says—asking what legitimate purposes there are *for the concept*—without some way of fixing the concept, the language, or the practice?

In comparison, conventional revisionist accounts have a ready answer. Revisionists have typically started with some diagnostic story about our

responsibility-characteristic practices, attitudes, and judgments. This does not mean an abandonment of the ameliorative concern. Indeed, such concerns are exactly why revisionists can maintain that when we encounter difficulties with the folk notion it makes sense to ask the questions that figure in Holroyd's construal of ameliorative theorizing—we can ask what work the concept does for us, what work we would like it to do in light of what we think about the associated practices, attitudes, and judgments, and we then ask whether there is a normatively adequate thing that does *that*—even if it departs from folk usage.[2]

Put this way, there is not a lot of daylight between revisionists and ameliorative approaches. Perhaps the ameliorative theorist can insist that she, unlike the revisionist, is willing to walk away from concerns about *moral responsibility* in the interest of responsibility-like practices that are not responsibility. As we have seen above, this misses the existence of "replacementism" or denotational versions of revisionism. Moreover, it also raises for itself challenges that revisionists have been at pains to respond to. That is, ameliorative theories that readily dispense with any connection to our responsibility-characteristic attitudes, practices, and judgments run the risk of theft over honest toil if they claim to be a theory of some *x* but simultaneously grant that it might not be about *x* at all. The interest in an ameliorative theory of *moral responsibility* (but in the case under consideration, strictly speaking, not moral responsibility) looks like it is parasitic on the theory having some recognizable connection to moral responsibility. So, rather than eschewing the resources of revisionism, it seems more promising to think of ameliorative theorizing as a species of revisionism, one that is accepted by many current revisionists.

Desert

Here is a view that has been expressed by a number of philosophers: although there may be multiple senses of responsibility, the one that centrally figures in debates about moral responsibility necessarily involves the idea of moral desert (Pereboom, 2001; Pereboom, 2014; Pereboom, 2017; Caruso and Morris, 2017). If this is right, though, it suggests a serious difficulty for moral influence or instrumentalist views of moral responsibility. The worry is that such views lack the resources for capturing the operative sense of desert. Given that the view in Vargas (2013) is instrumentalist in this way, it seems vulnerable to the objection that it is not responsive to the core issue in the philosophical debate about moral responsibility, on grounds that it cannot capture the operative notion of basic desert.

The most influential version of this position has been articulated by Derk Pereboom in a number of different places (Cf. Pereboom, 2014, p. 2; Pereboom, 2017, p. 260). On his account, the conception of desert at stake in debates about moral responsibility (and free will) is something he calls *basic desert*. The idea

of basic desert is that an agent "would deserve to be blamed or praised just because she has performed the action, given an understanding of its moral status, and not, for example, merely by virtue of consequentialist or contractualist considerations" (2014, p. 2; cf. Doris, 2015). In virtue of ruling out consequentialist considerations, basic desert seems incompatible with instrumentalist approaches to responsibility. So, if basic desert captures a core feature of the notion of responsibility at stake in debates about responsibility, then instrumentalist theories fail to be theories about moral responsibility.

As noted above, one possible response on behalf of the revisionist is to simply accept some potentially significant revisionism about desert, and thus, about responsibility. This answer would come at the cost of making the revisionist proposal not about moral responsibility, but instead, about some nearby notion. If theoretical push becomes normative shove, then this is exactly what the revisionist can do, going in for a form of replacementism or denotational revisionism. It would mean that the account would not enjoy a ready explanation of its significance, but the instrumentalism would also provide an answer about why this not-quite-responsibility notion has normative significance for us.

Three other possibilities are available, though. First, instrumentalists (and others) could reject the idea that desert is essential to moral responsibility. Michael McKenna (2012) has explored a version of this view. Ultimately, he argues that a desert-invoking formulation is compatible with his communicative account of responsibility. In a related but distinct vein, David Shoemaker (2015) has suggested that questions of desert are restricted to harsh treatment and not central to responsibility. Apart from these important exceptions, comparatively few contemporary accounts of responsibility have explicitly repudiated the centrality of desert.

If the presumption that desert is central to moral responsibility is widespread—and it seems to be—a more satisfying answer from the moral influence theorist needs to be formulated. There is a further consideration here. Recently, Dana Nelkin (2016) has argued that desert and accountability are mutually entailing. If that is right, then to the extent that going instrumentalist theories are intended as accounts of accountability, they must take on board some notion of desert. So, the most obvious alternatives for the instrumentalist are these: accept the basic desert requirement and argue that it can be met, or maintain that some notion of desert is involved in responsibility, but not basic desert.

How might these responses go? Some context is in order. On Pereboom's canonical statement of basic desert, the basis of desert is simply (1) the nature of the agent (a responsible agent) and (2) the moral quality of the action. As noted above, he emphasizes that the blaming cannot be licensed by consequentialist or contractualist considerations (2014, p. 2). One thing to note is that this characterization is supposed to be a neutral characterization of the notion of desert at stake in debates about moral responsibility. In principle, it is

something that all parties should be able to agree upon, even if they do not share Pereboom's commitment to the incompatibility of moral responsibility with determinism, his rejection of free will, and so on.[3]

Here is a promising analysis of the basis of desert in basic desert: desert is the product of two independent variables: culpability and wrongdoing (Brink, 2012, p. 498; Brink and Nelkin, 2013). This sort of analysis has structural similarities to Pereboom's account, in that there is an agent-focused element and an action-focused element.[4] It also provides a more specific elaboration of the basis of desert. If an instrumentalist account can capture this even narrower idea, it will have provided a conception of desert that satisfies Pereboom's notion of basic desert. The remainder of this section argues that, at least on the approach given at the outset of this paper, the idea of basic desert can be accommodated.

Recall that on the account given above—a two-tiered instrumentalist account—the responsibility-characteristic attitudes, practices, and judgments can be backward looking. Within the practice, whether one deserves blame is not settled by instrumentalist considerations. Instrumental considerations arise at the level of whether to have the practice at all. They need not arise at all within the practice. So, on a two-tiered theory, the propriety of first-order, substantive judgments of desert can be settled by culpability and wrongdoing—or, by the nature of the agent and the act. This structure mimics something of the flavor of Strawson's idea that there are questions internal to a practice and questions external to a practice, and that an adequate (perhaps consequentialist) explanation of why we have the practice is not yet an adequate explanation of the propriety conditions internal to the practice (Strawson, 1962; Strawson, 1985).

What the "no consequentialism" constraint of Pereboom's basic desert formulation rightly captures is the idea that there is a conception of desert on which consequences are entirely irrelevant. This non-consequentialist notion of desert is plausibly at play in our desert-entailing judgments of responsibility. As Doris (2015) has argued, trying to shoehorn in consequences into that notion of desert does considerable violence to this notion of desert. The two-tiered approach given above respects this constraint, and permits judgments of desert to be basic desert judgments. Internal to the practice—in typical first-order, substantive judgments about responsibility—the question of consequences does not come up at all.

Of course, two-tiered structures have had their critics. They have also had their proponents (Rawls, 1955; Hart, 1959; Copp, 1995; Hooker, 2000; Parfit, 2011; Miller, 2014b). There are familiar challenges about scapegoating, self-effacement, and the like. There are also established replies to these worries (Arneson, 2003; Miller, 2014c; Miller, 2014b; Vargas, forthcoming). The point here is not to litigate whether two-tiered approaches to a subset of the normative (i.e., to responsibility, as opposed to all of morality) is viable. The present ambi-

tion is only to show that the account given above can satisfy the putative requirement that an adequate theory of moral responsibility must capture Pereboomian basic desert.

What might Pereboom say to the foregoing? Pereboom recognizes that some instrumentalist accounts retain a notion of desert on consequentialist grounds (2017, p. 262). However, he suggests that this is an ersatz notion of desert. He goes on to say that what these accounts call "deserved responses are really just negative or positive incentives" (p. 262).

This reply, at least in this form, is not persuasive. Instrumentalist justifications of a practice, qua practice, need not infect the nature of first-order judgments internal to those practices, and this is something we readily recognize in contexts outside of responsibility. Again, consider the nature of foul calls.

Foul calls in sports are typically justified in light of the need to protect the players while preserving the enjoyment of the sport. This is a straightforwardly instrumental justification. Internal to a game, however, whether a foul is rightly or wrongly called is settled by the rules. In a given instance, the proper application of these rules can and does come apart from the justification for having those rules and not some other. When a foul is called, it is deserved or not in light of what happens in the game. This is true whether or not the application of the rules does anything for the safety of the involved players. It can also be true that a foul can be deserved even when it makes the game less enjoyable for athlete and spectator alike. Fans can and would get righteously angry about an undeserved foul call in the last game ever played of any sports game, even when there is no sense left to be made of the idea of an incentive. To suggest that a rightly-called foul is "really just a safety incentive" (or what have you) runs together what a foul call *is* with what *justifies* our having foul calls (Vargas 2015a). So, it seems the account of responsibility proposed above can accommodate both the idea that responsibility attributions must be desert-entailing, and the idea that these judgments must be of the basic desert variety.

Before concluding, it is worth noting that there is a further approach available to those who find appealing the present approach to responsibility, but worry about accommodating desert judgments. One might accept that an adequate theory of responsibility must capture desert while rejecting the idea that what must be captured is *basic* desert. There are various routes here, but the following is one that fits with the methodological commitments given above. It proceeds by indicting the basis for thinking that basic desert is central to an adequate theory of moral responsibility.

Recall the difference between conceptualist and phenomenalist construals of the philosophical stakes of a theory of responsibility. Pereboom, Caruso, and other proponents of basic desert are readily interpreted as conceptualists, that is, as methodologically committed to capturing our concepts, thoughts, or beliefs about the term "responsibility." However, by phenomenalist lights,

this is mostly a mistake, or more charitably, something mostly peripheral to an adequate theory of moral responsibility. On the phenomenalist reading—a reading convergent with the ameliorative impulses that animate Holroyd, for example—the proper philosophical stakes are about the nature and normative integrity of our practices. *These* practices, the phenomenalist might say (while gesturing at various phenomena in our social world), and not some armchair stipulation of a metaphysics of desert, are the subject of a theory of responsibility. These things—the phenomena of the world—are plausibly the subject of most accounts in the long history of philosophical debates about responsibility. If it turns out that what we need to explain, which actual or possible practices are normatively appealing, which attitudes and judgments are justified and have normative force, is some non-basic notion of desert, then so much the worse for basic desert.

The conceptualist might protest that this makes compatibilist theories of responsibility too easy, or that it eliminates any substantive difference between compatibilism and hard incompatibilism. This thought is in the spirit of some remarks made by Pereboom (2014, pp. 2–3; 2017, p. 260). Notice, though, that nothing in the phenomenalist construal of the subject matter rules out the possibility that our ordinary practices may presume that we have impossible or unlikely forms of agency, or that our practices are normatively indefensible. Incompatibilism and responsibility eliminativism remain live options. That ordinary convictions invoke demanding forms of agency is just what is claimed by many incompatibilists and revisionists. The conviction that something like our practices of moral blame are indefensible is what separates revisionists from eliminativist or hard incompatibilists.

Once we abandon fights about our favorite armchair formulae of folk beliefs and focus on the phenomena of our moral life and the best ways to rationally and normatively parse those things, we may find that revisionist, phenomenalist approaches to responsibility may be what many parties—hard incompatibilist and conventional compatibilists alike—have been searching for all along.

Notes

1. One interpretive issue concerns how much Holroyd is inclined to read the ameliorative analysis as a competitor to revisionist approaches as opposed to a species of revisionism. The tenor of her discussion strongly suggests that she has the latter in mind (otherwise, her focus on the putative distinction between the approaches is harder to make sense). However, the initial framing of her pursuing a project of "an explicitly ameliorative analysis of responsibility, rather than as merely a revisionary project" (156–7) could be read as signaling openness to a *species/genus* reading, rather than a *competitor* reading. The present discussion assumes a competitor reading, but some of the objections raised below concerning the lack of daylight between Holroyd's ameliorative

approach and some standard revisionist approaches would also serve, with only min modifications, as objections to the species/genus reading.

2. Haslanger (2012) expresses some skepticism about there always being *the* co cept of X, as there can be different things that fit that characterization—the tacit con mitments, the operative concept, the cluster of genealogical/social/ideological thin; that gave rise to that concept. The view in Vargas (2013) is compatible with this. Indee in that work and other places, the view is frequently explicit that there may be mar concepts of responsibility, as well as several ways of understanding what concepts *sim pliciter* are. For Haslanger, her characterization of this terrain is partly motivated by th possibility of various kinds of externalism about concepts—and this too figures in th motivation for the revisionism on offer in *Building Better Beings* (Vargas, 2013).

3. Notice that one could reject the idea that we can give a partisan-neutral char acterization of desert. Galen Strawson's (1994) account of responsibility nihilism depend on view of desert that could make sense of, for example, eternal damnation and eterna reward, and this is arguably more than most compatibilist accounts hope to deliver However, Pereboom's construal of basic desert is intended as a neutral proposal abou the stakes of the incompatibility debate, so its formulation cannot involve so demanding a conception of desert.

4. This account provides a compatibilist-friendly (where compatibilism is the view that free will and moral responsibility are compatible with determinism) way to capture the idea that punishment is properly proportional to desert. For those who maintain that the stakes of the responsibility debate require a notion of desert or responsibility that supports retributivism (cf. Caruso and Morris 2017), this is one way compatibilists might meet that demand. Of course, a number of compatibilists and ethicists more generally have been skeptical of retributivism, and thus, would reject as misguided demands that make retribution-supporting notions of responsibility the central issue (cf. Scanlon, 1988; Wallace, 1994).

References

Arneson, R.J. (2003). The Smart Theory of Moral Responsibility and Desert. In S. Olsaretti (ed.), *Desert and Justice* (pp. 233–258). Oxford: Oxford University Press.

Bennett, C. (2002). The Varieties of Retributive Experience. *The Philosophical Quarterly*, 52(207), 145–163.

Brink, D.O. (2012). Retributivism and Legal Moralism. *Ratio Juris*, 25(4), 496–512.

Brink, D.O., and Nelkin, D. (2013). Fairness and the Architecture of Responsibility. *Oxford Studies in Agency and Responsibility*, 1, 284–314.

Caruso, G.D., and Morris, S.G. (2017). Compatibilism and Retributive Desert Moral Responsibility: On What is of Central Philosophical and Practical Importance. *Erkenntnis*, 82(4), 837–855.

Copp, D. (1995). *Morality, Normativity, and Society*. New York: Oxford University Press.

Doris, J. (2015). Doing Without (Arguing about) Desert. *Philosophical Studies*, 172(10), 2625–2634.

Fischer, J.M., and Ravizza, M. (1998). *Responsibility and Control: A Theory of Moral Responsibility*. New York: Cambridge University Press.

Fricker, M. (2016). What's the Point of Blame? A Paradigm Based Explanation. *Nous*, 50(1), 165–183.

Hart, H.L.A. (1959). Prolegomena to the principles of punishment. *Proceedings of the Aristotelian Society New Series*, 60, 1–26.

Haslanger, S. (2012). *Resisting Reality: Social Construction and Social Critique*. New York: Oxford University Press.

Holroyd, J. (2018). Two Ways of Socializing Moral Responsibility: Circumstantialism versus Scaffolded-Responsiveness. In K. Hutchison, C. Mackenzie, and M. Oshana (eds.), *Social Dimensions of Moral Responsibility* (pp. 137–162). New York: Oxford University Press.

Hooker, B. (2000). *Ideal code, real world: A rule-consequentialist theory of morality*. New York: Clarendon Press.

Hurley, S. (2000). Is Responsibility Essentially Impossible? *Philosophical Studies*, 99, 229–268.

McCormick, K. (2013). Anchoring a revisionist account of moral responsibility. *Journal of Ethics and Social Philosophy*, 7(3), 1–19.

McCormick, K. (2015). Companions in Innocence: Defending a New Methodological Assumption About Moral Responsibility. *Philosophical Studies*, 172(2), 515–533.

McCormick, K. (2016). Revisionism. In K. Timpe, M. Griffith, and N. Levy (eds.), *Routledge Companion to Free Will* (pp. 109–120). New York: Routledge.

McCormick, K. (2017). Why We Should(n't) Be Discretionists About Free Will. *Philosophical Studies*, 174(10), 2489–2498.

McCormick, K. (forthcoming). Meeting the Eliminativist Burden. *Social Philosophy and Policy*, 36(1).

McGeer, V. (2014). P.F. Strawson's Consequentialism. In D. Shoemaker and N.A. Tognazzini (eds.), *Oxford Studies in Agency and Responsibility Volume 2* (pp. 64–92). New York: Oxford University Press.

McGeer, V. (2015). Building a Better Theory of Responsibility. *Philosophical Studies*, 172(10), 2635–2649.

McGeer, V. (forthcoming). Scaffolding Agency: A Proleptic Account of the Reactive Attitudes. *European Journal of Philosophy*.

McKenna, M. (2012). *Conversation and Responsibility*. New York: Oxford University Press.

McKenna, M. (2013). Reasons-Responsiveness, Agents, and Mechanisms. In D. Shoemaker (ed.), *Oxford Studies in Agency and Responsibility 1* (pp. 151–184). New York: Oxford University Press.

Miller, D.E. (2014a). "Freedom and Resentment" and Consequentialism: Why "Strawson's Point" Is Not Strawson's Point. *Journal of Ethics and Social Philosophy*, 8(2), 1–22.

Miller, D.E. (2014b). Rule Utilitarianism. In B. Eggleston and D.E. Miller (eds.), *The Cambridge Companion to Utilitarianism* (pp. 146–165). Cambridge: Cambridge University Press.

Miller, D.E. (2014c). Reactive Attitudes and the Hare-Williams Debate: Towards a New Consequentialist Moral Psychology. *The Philosophical Quarterly*, 64(254), 39–59.

Nelkin, D. (2016). Accountability and Desert. *Journal of Ethics*, 20(1–3), 173–189.

Nelkin, D.K. (2011). *Making Sense of Freedom and Responsibility*. Oxford: Oxford University Press.

Nichols, S. (2006). Folk Intuitions on Free Will. *Journal of Cognition and Culture*, 6(1 & 2), 57–86.

Nichols, S. (2015). *Bound: Essays on Free Will and Responsibility*. New York: Oxford University Press.

Nichols, S. (2017). Replies to Kane, McCormick, and Vargas. *Philosophical Studies*, 174(10), 2511–2523.

Nowell-Smith, P. (1948). Free Will and Moral Responsibility. *Mind*, 57(225), 45–61.

Parfit, D. (2011). *On What Matters* (1). New York: Oxford University Press.

Pereboom, D. (2001). *Living Without Free Will*. Cambridge: Cambridge University Press.

Pereboom, D. (2014). *Free will, agency, and meaning in life*. New York: Oxford University Press.

Pereboom, D. (2017). Response to Daniel Dennett on Free Will Skepticism. *Rivista Internazionale di Filosofia e Psicologia*, 8(3), 259–265.

Rawls, J. (1955). Two Concepts of Rules. *Philosophical Review*, 64, 3–32.

Raz, J. (2011). *From Normativity to Responsibility*. Oxford: Oxford University Press.

Scanlon, T.M. (1988). The Significance of Choice. In S.M. McMurrin (ed.), *The Tanner Lectures on Human Values* (pp. 150–216). Cambridge: Cambridge University Press.

Shoemaker, D. (2015). *Responsibility from the Margins*. New York: Oxford University Press.

Singer, I. (2002). Freedom and Revision. *Southwest Philosophy Review*, 18(2), 25–44.

Smart, J.J.C. (1961). Free Will, Praise, and Blame. *Mind*, 70, 291–306.

Strawson, G. (1994). The Impossibility of Moral Responsibility. *Philosophical Studies*, 75, 5–24.

Strawson, P.F. (1962). Freedom and Resentment. *Proceedings of the British Academy* XLVIII, 1–25.

Strawson, P.F. (1985). *Skepticism and naturalism: some varieties*. New York: Columbia University Press.

Vargas, M. (2004). Responsibility and the Aims of Theory: Strawson and Revisionism. *Pacific Philosophical Quarterly*, 85(2), 218–241.

Vargas, M. (2011a). Revisionist Accounts of Free Will: Origins, Varieties, and Challenges. In R. Kane (ed.), *Oxford Handbook on Free Will, 2e* (pp. 457–484). New York: Oxford University Press.

Vargas, M. (2011b). The Revisionist Turn: Reflection on the Recent History of Work on Free Will. In J. Aguilar, A. Buckareff, and K. Frankish (eds.), *New Waves in the Philosophy of Action* (pp. 143–172). New York: Palgrave Macmillan.

Vargas, M. (2013). *Building Better Beings: A Theory of Moral Responsibility*. Oxford, U.K.: Oxford University Press.

Vargas, M. (forthcoming). Instrumentalist Theories of Moral Responsibility. In D. Nelkin and D. Pereboom (eds.), *Oxford Handbook on Moral Responsibility*. New York: Oxford University Press.

Vargas, M. (forthcoming). Revisionism. In J. Campbell (ed.), *A Companion to Free Will*. Oxford: Wiley-Blackwell.

Vargas, M.R. (2015a). Desert, Responsibility, and Justification: Reply to Doris, McGeer, and Robinson. *Philosophical Studies*, 172(10), 2659–2678.

Vargas, M.R. (2015b). Précis of *Building Better Beings: A Theory of Moral Responsibility*. *Philosophical Studies*, 172(10), 2621–2623.

Wallace, R. J. (1994). *Responsibility and the Moral Sentiments*. Cambridge, MA: Harvard University Press.

Wolf, S. (1990). *Freedom Within Reason*. New York: Oxford University Press.

Manuel Vargas is a professor of philosophy at the University of California, San Diego. He is the author of *Building Better Beings: A Theory of Moral Responsibility*, and a co-author of *Four Views on Free Will*. UC San Diego Philosophy Dept., 9500 Gilman Drive #0119, La Jolla, CA 92093-0119. <mrvargas@ucsd.edu>.