

University of Pennsylvania Carey Law School

Penn Law: Legal Scholarship Repository

Faculty Scholarship at Penn Law

9-26-2021

Blameworthiness, Desert, and Luck

Mitchell N. Berman

University of Pennsylvania Carey Law School

Follow this and additional works at: https://scholarship.law.upenn.edu/faculty_scholarship



Part of the [Criminal Law Commons](#), [Criminology and Criminal Justice Commons](#), [Ethics and Political Philosophy Commons](#), [Law and Philosophy Commons](#), [Law Enforcement and Corrections Commons](#), and the [Public Law and Legal Theory Commons](#)

Repository Citation

Berman, Mitchell N., "Blameworthiness, Desert, and Luck" (2021). *Faculty Scholarship at Penn Law*. 2801. https://scholarship.law.upenn.edu/faculty_scholarship/2801

This Article is brought to you for free and open access by Penn Law: Legal Scholarship Repository. It has been accepted for inclusion in Faculty Scholarship at Penn Law by an authorized administrator of Penn Law: Legal Scholarship Repository. For more information, please contact PennlawIR@law.upenn.edu.

Forthcoming: *Noûs* (2022)

Blameworthiness, Desert, and Luck

Mitchell N. Berman*

Draft, September 26, 2021

Abstract: Philosophers disagree about whether outcome luck can affect an agent's "moral responsibility." Focusing on responsibility's "negative side," some maintain, and others deny, that an action's results bear constitutively on how "blameworthy" the actor is, and on how much blame or punishment they "deserve." Crucially, both sides to the debate assume that an actor's blameworthiness and negative desert are equally affected—or unaffected—by an action's results. This article challenges that previously overlooked assumption, arguing that blameworthiness and desert are distinct moral notions that serve distinct normative functions: blameworthiness serves a *liability* function (removing a bar to otherwise impermissible treatments), whereas desert serves a *favoring* function (contributing new value to states of affairs, or providing new reasons for responsive treatments). Having distinguished (negative) desert from blameworthiness, the article proposes a novel resolution to the outcome-luck debate: that results do not affect an agent's liability to blame, but do affect the amount and severity of blame to which the agent is justly liable, including by affecting the severity of blame that the agent deserves.

Introduction

A standard case in the moral luck literature takes the following form: H and M, acting independently, both shoot at V, intending to kill, without justification or excuse, and (if it matters) while holding equal credences regarding the likelihood of succeeding. Due to factors outside of the control or anticipation of H or M (a gust of wind, an imperfection in M's rifle barrel, V's sudden movement), H's bullet hits V, killing them, while M's bullet flies harmlessly past. (H hit V; M missed V.)

Philosophers use hypotheticals like this one—or others that involve paired agents who act recklessly or negligently with respect to the result at

issue, rather than intentionally¹ —to motivate questions about moral outcome luck that they frequently formulate (when not deploying that frustratingly plastic word “responsibility”) in terms of “blameworthiness” or “desert.” They ask, for example, whether H and M are “equally blameworthy,”² or whether H and M “deserve equal punishment.”³ Frequently, commentators speak both ways at once. To pick a single example almost at random, Susan Wolf, in a well-known article (2001, 5-6, emphases added), characterizes the outcome luck question, interchangeably, as “whether those whose acts actually lead to serious harm *deserve* the same treatment . . . as those who, but for fortune, would have caused as much damage” and as whether “a person whose actions have morally worse effects is . . . more *blameworthy* than one whose equally faulty behavior has less harmful consequences.” Similarly, the most prominent living retributivist, Michael Moore (1997, 215, 218) poses the “question” and “problem” of “moral luck” in terms, interchangeably, of “blameworthiness,” “moral deserts,” and “moral responsibility.”

The assumption that underlies this pattern of speech, occasionally made explicit,⁴ is that to be blameworthy for thus-and-such and to deserve punishment (or suffering, or hard treatment) for thus-and-such are linked

* Leon Meltzer Professor of Law, the University of Pennsylvania Carey Law School, and Professor of Philosophy, the University of Pennsylvania. Email: mitchberman@law.upenn.edu. Near and distant predecessors to this paper were presented at Yale Law School, the Centre for Ethics, Law and Public Affairs at the University of Warwick, and at the APA Eastern. I am grateful to workshop participants, especially Randolph Clarke, Matthew Clayton, Paulina Essunger, Helen McCabe, Tom Parr, Scott Shapiro, Adam Slavny, Victor Tadros, and Gideon Yaffe, for constructive feedback on these occasions. I am also indebted to Craig Agule, David Enoch, Dana Nelkin, Michael Zimmerman, and two anonymous referees for very helpful written comments, and to Gideon Rosen for graciously welcoming me to audit his course on moral responsibility.

¹ I highlight cases in which the results are intended because, like others (Sverdlik 1988, 79), I find them simpler. But, in broad strokes, my analysis applies the same to cases of intent, knowledge, foresight, and negligence. The analyses might differ at the level of detail, and I identify some possible differences in due course.

² Examples of formulations along these lines include, among countless others, Edwards & Simester (2019, 65); Nelkin (2019); Hartman (2019, 3181), Otsuka (2009, 373); Enoch & Marmor (2008, 409); Domskey (2004, 445); Zimmerman (2002, 561); Sverdlik (1988, 79).

³ Again, just a sample of numerous instances are Freiman & Nichols (2011, 124); Levy (2005, 265-66); Christopher (2004, 421); Rosebury (1995, 499); Lewis (1989, 55-56).

⁴ For example, Andrew Khoury (2018, 781) squarely maintains that the puzzle of resultant moral luck involves the extent of an agent’s blameworthiness or praiseworthiness and then reasons: “It is natural to think, from the retributivist standpoint, that just punishment is a matter of giving the criminal what he deserves and that what he deserves is proportionate to his blameworthiness. If the successful are more blameworthy than the unsuccessful, then they deserve harsher punishment.”

in such a fashion that questions formulated in terms of blameworthiness and deserved punishment necessarily elicit the same answers. Those who believe in outcome luck will conclude that H is more blameworthy than M and deserves more punishment or harsher treatment; those who reject outcome luck will conclude that H and M are equally blameworthy and deserve equal punishment. I will call the premise that the factors that determine the fact and extent of an agent's blameworthiness are identical to those that determine the fact and extent of an agent's negative desert *the equivalence thesis*, shorthand for "the blameworthiness-desert equivalence thesis."⁵ The equivalence thesis is nearly universal. Not only is it shared by proponents and opponents of moral outcome luck,⁶ it is endorsed or assumed by scholars writing on topics across the waterfront of punishment theory. But if the equivalence thesis is false, then these two questions about moral outcome luck—phrased in terms of "blameworthiness" and of "desert"—are not alternative formulations of the same substantive question, but different substantive questions that can (though might not) support different answers.

I argue in this paper that the equivalence thesis is false. I'd like to further conclude that one of these normative concepts is sensitive to the results of an agent's willings and that the other is not. That would furnish a neat and crisp resolution to the outcome-luck debates. Unfortunately, matters prove to be more complex and equivocal. They are equivocal because whether negative desert is sensitive to outcomes depends upon an account of the point of or rationale for (negative) desert in the first place, an issue that this article hopes to advance but does not purport to resolve. They are complex because, just as negative desert must be distinguished from blameworthiness, a second largely overlooked distinction must be drawn internal to blameworthiness, the distinction between the extent to which an agent is liable to blame for their conduct, and the magnitude or severity of blame to which they are liable.

⁵ People who harbor this assumption are overwhelmingly likely to believe that equivalence holds as well between praiseworthiness and deserved reward. But to simplify discussion, and because it is the clear focal case in the literature, I will concentrate on the "negative" side—the side that involves blameworthiness and deserved punishment—not the "positive" side that involves praiseworthiness and deserved reward. Readers who deny that anybody ever deserves bads (to be punished, to suffer, etc.) but accept that persons can deserve goods (to be rewarded, to be happy, etc.) can follow along by making the necessary substitutions in what follows.

⁶ The only counterexamples I'm aware of are persons who combine belief in moral outcome luck with the denial of desert, either negative or total. Acceptance of moral outcome luck leads to the conclusion that H is more blameworthy than M. Denial of desert entails the conclusion that H and M deserve the same punishment, namely none.

The paper proceeds over five sections. Section 1 introduces two distinct functions that responsibility concepts are needed to serve: liability and favoring. Sections 2 and 3 argue, respectively, that desert is a favoring concept, and that blameworthiness serves a liability function. Section 4 examines how the results of an agent's conduct bear on their desert. It explains that an answer to that question depends upon an answer to the logically prior question of why the fact of wrongdoing, results aside, would favor that the wrongdoer suffer or be punished or blamed in the first place. That is, we cannot settle on confident judgments about the grounds or determinants of deserved blame or suffering—and, in particular, whether those grounds or determinants include results of an agent's willings—without a clearer grasp of what justifies our embrace of negative desert as a moral concept that serves a favoring function. Developing and defending an account of desert's point or function is a large undertaking that cannot be attempted here. But Section 4 presents the skeleton of one account supportive of the conclusion that desert is sensitive to (some) results. Section 5 examines how the results of an agent's conduct bear on their blameworthiness. It argues that results matter—there is outcome luck—in regards to how much an agent is liable to blame, but not in regards to how much blame an agent is liable to.

1. Two Responsibility Concepts: Liability and Favoring

Two distinct paths into our topic immediately suggest themselves. The more obvious is to start with our concepts desert and blameworthiness, aiming to show that, properly analyzed, they serve different normative functions and rest on different sets of determinants. A second would start by identifying the normative functions that we need our moral responsibility concepts to serve, aiming to show that desert and blameworthiness are best assigned different functions (where the relevant criteria include fit with current usage and theoretical utility). If my substantive claims about desert and blameworthiness are correct and not highly revisionary then both routes—concepts first, or functions first—should be available and each should have the capacity to deliver us to the same place.

For two reasons, I will pursue the second approach. First, I'm betting that more readers will assent more quickly to what I claim about the functions that our responsibility concepts must serve than to my—or any—proposed analyses of the concepts. In particular, the concepts-first approach must wend its way through the fast-growing and markedly

fragmented literature on blameworthiness that the functions-first approach might render at least partially unnecessary.

Second, the functions-first approach more closely tracks the analytic framework that Gideon Rosen (2015) pressed in the course of presenting and defending his own analysis of blameworthiness (his “alethic conception”). I find that framework incisive and congenial. But I also think that its value is increased, not diminished, if we start, as Rosen himself does not, by attending carefully to distinct functions that our responsibility talk and practices serve. Proceeding functions-first allows me to make better use of Rosen’s argument, both as guide and as foil.

* * *

“A theory of moral responsibility,” Rosen begins, “should address three questions” (2015, 65):

The analytic question: What is it for a person to be morally responsible for an act, event, or state?

The grounding question: What are the conditions [and determinants] under which a person is morally responsible for an act, etc.?⁷

The explanatory question: Why are the conditions [and determinants] of moral responsibility as they are?

In addressing *the analytic question*, Rosen observes: “the phrase ‘moral responsibility’ has no clear sense—not in ordinary English, and not in philosophy.” Therefore, he says (2015, 66), “before we can raise substantive questions about the nature of his thing, we need to fix—by fiat—a definite meaning for the phrase.”⁸

I find Rosen’s parsing of the separate questions extremely helpful. I too will begin by addressing the analytic question and will turn later (sections 4 and 5) to the grounding and explanatory questions. But where Rosen infers from the fact that “moral responsibility” bears multiple meanings that we should stipulate a single answer to the analytic question, I propose to investigate the variety of questions to which our responsibility talk is

⁷ I take the inserted bracketed language to be a friendly amendment. See Rosen (2015, 65 n.1).

⁸ Rosen’s own proposal, which will not further concern us, is as follows: (1) A person is morally responsible if blameworthy or praiseworthy. (2) X is blameworthy for φ iff X merits moral blame for φ ; (3) X merits moral blame for φ if it would be appropriate to blame X for φ ; (4) It would be appropriate to blame X for φ iff the hostile reactive emotions toward X would be appropriate. According to Rosen’s proposed alethic conception, the appropriateness of reactive emotions depends upon the truth of the belief-like thoughts implicit in them.

responsive. Proceeding in this way may yield a non-arbitrary, non-stipulative multiplicity of senses of moral responsibility to analyze.

The first thing to note, as others have emphasized (e.g., Zimmerman 2015; Moore 1997, 197), is that responsibility talk of the sort that the moral luck discussion implicates arises in prototypically backward-looking contexts. This is not true of all responsibility talk. Some responsibility questions arise prospectively. We might say that only responsible agents should be permitted to enter into binding contracts. Or we may ask about the responsibilities that come with a particular role. But the natural contexts for debates over outcome luck are retrospective.

Think of morality as a system designed, first and foremost, to establish forward-looking norms regarding what we may, or must, do.⁹ At time t_n it addresses agent A, enjoining them not to ϕ . Suppose that A flouts that injunction. What are the moral consequences, at t_{n+x} , of A's norm-violation? How, if at all, has A's (wrongful) ϕ ing altered the moral landscape, as far as A is concerned? How, if at all, does A's "moral profile" (Greenberg 2014, 1308) change in virtue of A's ϕ ing?

Here are three possibilities. The first two are uncontroversial; the third is disputed.

First, in virtue of A's ϕ ing, A sometimes assumes or acquires *new moral obligations*—what W.D. Ross (1930, 21) termed "duties of reparation" and others call "residual duties."¹⁰ A must apologize to those they have wronged, and try to repair any damage. Scholars have argued, for example, that A must reorganize their life to minimize the risk of reoffending (Lee 2009), or must provide their victims with protection against wrongdoing by others (Tadros 2011). Second, A sometimes becomes *liable* to treatments by others that would be unfair to A, or would otherwise treat A wrongfully, but for A's ϕ ing. Third, in addition to making A liable to treatments to which A had been immune, A's ϕ ing could create new *reasons* for some agent or class of agents to subject A to such treatments, or could alter the non-instrumental impersonal *value* of states of affairs that involve A. It could be, for example, that the state of affairs in which A experiences suffering or anguish would have been bad absent A's ϕ ing but now becomes good. In sum, A's wrongdoing can possibly ground at least three different, but not incompatible, moral consequences: A's ϕ ing grounds or activates duties imposed on A; A's ϕ ing grounds liability to treatment that,

⁹ The notion of "design" here is functional and evolutionary, not agentially directed.

¹⁰ Philosophers debate whether A's ϕ ing grounds a new duty or activates the antecedent of an existing conditional duty to which A was already subject. For my limited purposes, I believe and hope that we needn't choose between wide-scope and narrow-scope construals of the relevant duties.

otherwise, would be unfair to A or would wrong A; A's ϕ ing grounds reasons to treat A in ways that would not otherwise obtain.

Three points need be made about these three potential changes to A's moral profile triggered by A's earlier conduct. First, none of the three changes is a bare "assessment" or "appraisal." In his classic essay, "Moral Luck," Thomas Nagel (1979, ch.3) repeatedly formulates the anti-luck position as insisting that "moral assessments" of persons should not be affected by factors outside their control. But as many commentators have noted (e.g., Nelkin 2019), it's not clear what assessments and appraisals are or involve. On the dominant interpretation, they are merely evaluations of the actor's character—judgments about whether or to what extent the actor is a "bad person" (Thomson 1989, 210 & n.3), or possesses or lacks "moral worth." (Greco 1995, 83). Construed that way, it is generally accepted that moral assessments are insensitive to outcome luck: at least as far as these paired episodes reveal, H and M are equally bad people and have equally degraded moral worth. Be that as it may, if these assessments are themselves morally inert—if they possess no further moral force—they do not serve a normative function and are not my concern here. Contrariwise, if different moral assessments of an agent do automatically trigger different moral consequences, then the task is to focus on those moral consequences, not on the assessments.

Second, the conditions that trigger the three potential changes might differ. Many people are apt to conclude that the conditions that trigger residual duties are less stringent than those that trigger liability, which are in turn less stringent than those that make a harsh response favored.¹¹

Third, all three types of changes implicate what are fairly conceived of as kinds of "moral responsibility." Inquiries into whether A shoulders a *residual duty*, or is *liable to* blaming treatments, or *should face* unwelcome consequences, on account of A's ϕ ing, could all be couched in terms of whether A was, in the relevant sense, "responsible" for ϕ ing. Still, if our actions can produce these significant changes to our normative profiles, and if the conditions under which the changes occur are at all regular (as they should be), then it is reasonable to expect the normative community to have formulated more particularized responsibility vocabulary that fits and marks the conditions and relationships at least tolerably well. Those expectations could be disappointed, but that would be surprising as well as unhappy.

¹¹ Rosen (2015, 66) seems to agree with the first contrast; having overlooked the difference between liability and favoring functions, he cannot express a view about the second.

2. Desert as Favored

It is a commonplace that desert is among the most mysterious and undertheorized of our normative concepts. Analysis of desert claims frequently starts with Joel Feinberg's (1970, ch.4) tripartite framework according to which an agent deserves some treatment or state of affairs (the "desert object") on account of some quality or action of the agent (the "desert base"). Feinberg's analysis is justifiably recognized as a significant advance. But it was advanced (as "Justice and Personal Desert") in 1963. Our shared understanding of general features of desert claims has not significantly advanced in the half-century since. Victor Tadros (2017, 394) fairly captured the state of our understanding of desert when wryly observing, in a review of Shelly Kagan's 2012 book *The Geometry of Desert*, that work on the topic is "so underdeveloped" that "the mere 24 footnotes and 7 references" in Kagan's 656-page tome "does not imply a failure to acknowledge the achievements of others writing in the field."

Some contributions to a smallish literature investigate features of particular desert relationships. Philosophers of criminal law debate just what it is (if anything) that wrongdoers deserve (to suffer, to be punished, to be censured, etc.),¹² and what conditions in addition to, or in lieu of, wrongdoing (e.g., culpability, harm) must obtain for that desert object to be deserved. Philosophers interested in distributive justice explore the bases for deserved economic resources. But the question that concerns us here is about desert in general. It is a matter of desert's "normative force" or "normative significance."

Suppose the following claims are true: A deserves a medal; B deserves to win the race; C deserves the job; D deserves to be blamed; E deserves to be happy; F deserves to be miserable; G deserves to be punished; H deserves the nomination. The question about desert's normative force asks what are the deontic or axiological entailments of the fact that an agent deserves some treatment or state of affairs, regardless of what that desert object, φ , happens to be.

Philosophers have offered, or assumed, two basic answers to this question. (e.g., Olsaretti 2003, 8; McKenna 2013, 135.) The first and older answer, chiefly associated with deontology, ties desert closely to duty and to justice: what it is for A to deserve φ on account of x is for justice to require

¹² Differences between "to suffer" and "to be punished" as retributivist desert objects are examined in Berman (2011).

that φ obtain for A, and for some class of agents to harbor a duty or obligation to bring φ about. The second answer, championed by consequentialists (e.g., Kagan 2012; Hurka 2001; Feldman 1997), holds that desert is a matter of value: what it is for A to deserve φ is for it to be impersonally, non-instrumentally good or valuable that φ be actualized. As one commentator (Kershnar 2010, xi) puts the latter view: “Desert is a relation that is an aspect of the good. Specifically, desert is a function that relates well-being, intrinsic value, and a ground. . . . Desert therefore looks at the explanation of what makes the world a better place.” Additional views combine these elements.¹³ What all the standard views of normative force have in common is a conception of desert as serving a favoring role, “either [as] as a matter of right or [as] a matter of (moral or nonmoral) value.” (Sher 1987, xi) It is testament to this truth that it was a struggle to introduce and explain the notion of favoring in section 1 without deploying the language of desert.

The fundamental debate in punishment theory between retributivist and consequentialist justifications for punishment reinforces that desert is a favoring concept. Despite many intramural disagreements within retributivism, common ground is that it justifies punishment “in terms of” the offender’s desert. (Berman 2016, 36-37). Here, the offender’s (supposed) desert plainly serves other than a liability function. It serves a favoring function. Retributivism is vigorously contested. Many philosophers think it barbaric. Why? Chiefly because the notion of desert it deploys is a favoring notion: A’s φ ing grounds an alteration in the impersonal noninstrumental value of A’s suffering (or being punished, or what-have-you), or grounds noninstrumental reasons for us to inflict punishment, or visit suffering, upon A.

¹³ For example, I have proposed that valid desert claims affect non-instrumental value and also confer weighty reasons or obligations, albeit conditionally. On my view (Berman 2013, 92), “That A deserves O on account of B means (a) given B, that A experience O (or that O obtain for A) is better than that A not experience O (or that O not obtain for A); and (b) if there is any agent or institution, X, with responsibility over the relevant domain, then X has a duty of justice to cause O to obtain for A.”

Fittingness conceptions of desert, according to which what it is for A to deserve φ (on account of x) is for it to be “fitting,” “apt,” or “appropriate” that A get φ , are not to the contrary. Fittingness conceptions, it seems to me, merely push back the question of desert’s normative force one step, pressing us now to inquire into the normative force or significance of the fact that it would be fitting or apt that A get φ . And the usual answers to that question associate duties of justice with duties to extend treatments that are fitting (because deserved), hence plainly treat desert as favoring. As Geoffrey Cupit (1996, 35) remarks: “To accept justice as fittingness is to accept that to treat people justly requires no more than that they be treated in accordance with their deserts.”

So we have a first-pass answer to the analytic question, with respect to *desert*:

The analytic answer (desert): What it is for an agent, A, to deserve a consequence or treatment, x , for an act, φ , is for it to be the case, in virtue of A's φ ing, that it is impersonally finally good for A to get x , or that some agent, B, has (weighty, special) reason to cause A to get x .

3. Blameworthiness as Liability

If desert captures the favoring notion of responsibility, it is natural to suppose that blameworthiness captures the liability notion. If for A to have negative *desert* for φ ing is for it to be the case that some negative or harsh treatment of A that would otherwise be disfavored or forbidden is now *avored*, then, very possibly, for A to be *blameworthy* for φ ing is for it to be the case that A is now *liable* to some negative or harsh treatment to which A would otherwise be immune. (This is not to say that the treatment at issue is rendered permissible *all things considered*; it's to say that a certain type of especially salient and forceful barrier to the treatment is erased, canceled, forfeited, or overridden.) Blameworthiness is a more modest upshot than desert, which is what debates over the justifiability of punishment would suggest: anti-retributivists will more readily accept that wrongdoing grounds blameworthiness—it better, else responsibility-based side constraints on instrumentalist goals would be in jeopardy—than that it grounds desert.

My proposal, then, is that to be blameworthy is to be liable to some otherwise proscribed “negative or harsh treatment.” It is meant to capture a common and intuitive view, but faces resistance from two (not incompatible) directions. It may be objected: first, that to be blameworthy is to be liable to *blame* (and not to some more vaguely or capaciously described forms of “harsh treatment”); and second, that to be blameworthy is to *deserve* blame or harsh treatment (and not merely to be “liable” to it).

Let's start with the first complaint, that my characterization of the responses to which the blameworthy agent is made liable is unnecessarily vague or objectionably general. The view I have in mind maintains that what being blameworthy renders one liable to is to being *blamed*. And, if so, we must carefully investigate what *blame* or *blaming* is to understand just what it is to be blameworthy. (See Graham 2014, 388-89) I think this doubtful even as a semantic matter. Were blameworthiness to have a narrower scope than I claim for it, one that licenses only those responses

that count as blame “in the strict sense,” then we should expect to have concept-words for related but distinct moral concepts: “castigation-worthy,” “punishment-worthy,” and so on. That (by and large) we don’t is evidence for the proposition that “blame” in “blameworthy” operates as a stand-in for a range of related practices. And this is just what ordinary lay discourse suggests: speakers claim that they or others are “blameless” or “not blameworthy” in speech acts intended to preserve the relevant actor’s immunity from a wide variety of (actual or putative) “blaming practices” — not only criminal punishment, but also vicarious liability in tort, military retaliation, a diminution of regard in the community, consumer boycotts, social stigma, rupture of friendships, and so on.¹⁴ In short, I think that we can fairly bypass the need to excavate “the essence of blame” (Coates and Tognazzini 2013, 8), by treating “blame” in “blameworthy” as shorthand for a variety of practices that share a family resemblance in involving directed criticism, censure, castigation, distancing, retaliation, infliction of costs or hardship, punishment, and the like.¹⁵

Now take the second basis for resisting my proposed answer to the analytic question as applied to blameworthiness: the common notion that “to be blameworthy is to deserve blame.” Of course, people use all these terms in diverse ways, and not everybody who says this intends to endorse a favoring conception of the word “deserve.” No doubt some instances of the utterance are meant to reflect the liability conception of

¹⁴ For a sampling of such varied claims, see Graeme Burton, “Morrisons tells court it’s ‘entirely blameless’ for 2014 payroll data leak: Firm claims it should not be held vicariously liable,” *The Inquirer*, Nov. 7, 2019; “Student killed in London knife attack ‘entirely blameless,’” *The Guardian*, Dec. 8, 2019; Omar Barghouti, “Boycott, Academic Freedom,, and the Moral Responsibility to Uphold Human Rights,” 4 *AAUP Journal of Academic Freedom* 6 (2013); “A Mom’s Perspective: Having a Child With an Eating Disorder,” <https://mirror-mirror.org/getting-help/a-moms-perspective>; Rick Smith, “Save Iowa’s 6,000 DACA Children,” *Iowa Starting Line*, Sept. 4, 2017, <https://iowastartingline.com/2017/09/04/save-iowas-6000-daca-children/>; “John and Munger suspended, Merwe withdraws appeal, no action on Tan,” *The New Paper*, Oct. 27, 2020.

¹⁵ Michael Zimmerman encourages this approach when observing that, “as Strawson himself notes, among the responses that are typically taken to be fitting to the quality of will manifested in someone’s behavior are responses that are not, or at least not merely, a matter of adopting some attitude but rather of engaging in certain more robust practices, such as the practices of reward and punishment.” (2015, 49). But he veers in a different direction when flirting with the possibility that “[i]f there are indefinitely many possible reactions, there may be indefinitely many kinds of responsibility.” (2015, 60). Zimmerman’s evident discomfort with that conclusion might suggest that kinds of retroactive moral responsibility are better individuated according to the broad kinds of changes to the moral profile distinguished in section 1 (activation of duties of repair, liability to responses, and favoring), rather than according to kinds of reactions to which one is liable. That approach does seem too granular for our shared lexical concepts.

blameworthiness, only with a different vocabulary.¹⁶ But many aren't. It seems plain that some authors who assert that to be blameworthy is to "deserve" blame thereby intend to assign blameworthiness a favoring function (perhaps in addition to a liability one).¹⁷ If desert serves a favoring function, and if part of what it is to be blameworthy is to deserve some response, then blameworthiness itself must be a favorer and not a bare matter of liability. Is it the case that for an agent to be blameworthy is for them to deserve harsh or unwelcome responses in the sense that some such responses directed to that agent are rendered not only permissible, but favored as well?

Again, I do not think so. And, again, I suspect that the contrary view depends overmuch on instincts about the semantics that do not withstand scrutiny. The starting assumption, Rosen says, is that an account of blameworthiness and praiseworthiness "should respect word structure." (2015, 66) "For X to be morally praise- or blameworthy for A is for X to be *worthy of*—or to *merit*—a certain type of response in light of A." And because "merit" is a near-synonym for "deserve,"¹⁸ the conclusion is hard to resist that to be blameworthy is to deserve blame.¹⁹

¹⁶ Some philosophers who say that for an agent to be blameworthy is for the agent to deserve blame, gloss what it is for A to "deserve blame" as it's being the case that A's being blamed is "warranted" (e.g., Arpaly 2003, 70-71). It is not always clear, though, whether such theorists treat an action's being warranted as a permissibility notion or a favoring one. As with "fitting" and "appropriate," (see note 12, above), "warranted" can be a waffle when not clearly defined or explicated in terms of such more familiar normative notions as "good," "reason," "duty," and "ought." This complaint has more force, however, insofar as our interest lies in first-order ethics (my concern in this paper) rather than in moral psychology.

¹⁷ I think this is true of Michael McKenna, for example, and in an illuminating way. Starting from a stated desire to accommodate, rather than to resist, "the familiar thought that blame is *deserved* by one who is blameworthy," (2013, 119) he ends up defending this striking principle of "directed blame": "Because it is a noninstrumental good that one who is blameworthy is harmed by the communicative practices constitutive of directed blaming, it is permissible to harm her by directly blaming her." (McKenna 2013, 136). I say the principle is striking because it is not generally true that a conclusion that some given conduct is *permissible* depends on the premise that it's *valuable*. So McKenna does not obviously need his controversial claim about noninstrumental positive value (that it is good to harm blameworthy agents by blaming them) to support his less controversial conclusion about permissibility. Possibly, then, McKenna has made his job harder than it need have been. If he has, that might be a consequence of his having associated the liability function that blameworthiness serves too closely to the favoring function that desert serves.

¹⁸ See Webster's Third New International Dictionary 1414 (offering one definition of "merit": to be or become deserving of good or ill).

¹⁹ Rosen does not himself take this last step, even though he does conclude—albeit tentatively and while granting its "weirdness" (Rosen 2015, 86)—that an agent is blameworthy if they deserve punishment, in the favoring sense. He reaches this conclusion on the basis of his surmise that one of the belief-like thoughts implicit in resentment is that

But resist it we should. We can accept the premise that an analysis of blameworthiness should (presumptively) respect word structure without accepting the conclusion that to be blameworthy is to “merit” or “deserve” blame, in the favoring sense. It is true that the suffix *-worthy* is sometimes defined as “deserving of.” But it can also mean “suitable or safe for.” That a vessel is seaworthy does not entail that it deserves to be put to sea in the sense that there is any intrinsic value in its being put to sea or that anyone has reason, simply in virtue of the vessel’s seaworthiness, to put it to sea. That a series is binge-worthy is not a reason to watch if you don’t watch tv, or lack a taste for shows in that genre. That one is trustworthy does not even entail that one deserves to be trusted; it merely signifies dependability, or that trusting would not be misguided. That an actor is trustworthy means only that *if* you have need for somebody to repose trust in (to give disinterested advice, to watch over your stuff), you wouldn’t be making a mistake to repose it in them.²⁰

Accordingly, for A to be blameworthy can mean only that at least some among a variety of ways of treating A that belong to the constellation of “blaming” behaviors—modes of treatment that would otherwise be unsuitable or inappropriate—are rendered suitable or apt. On this definition, it would not follow that there is any agent, B, of whom it is true, simply in virtue of the fact that A is blameworthy, that B has even a pro tanto reason to blame A, just as it is not the case that there is any agent, B, of whom it is true, simply in virtue of the fact that a craft is seaworthy, that B has a reason to put the craft to sea. In both cases, whether B should engage in the activity at issue (blaming A, putting the craft to sea) will depend upon the presence of reasons that are not grounded in the fact of A’s blameworthiness or seaworthiness, as the case may be.²¹

To be clear, my claim is not that the semantics require the interpretation I offer, but only that they don’t preclude it if other theoretical criteria

the object of one’s resentment should “suffer *in recognizing* what she has done.” (Rosen 2015, 82). I find what Rosen says about resentment persuasive. What remains unclear to me is, if agents can deserve “pain-in-recognition,” and if their deserving such pain is partly constitutive of their being blameworthy, what considerations should attract us, or do attract Rosen, to a Strawsonian rather than metaphysical account of blameworthiness in the first place.

²⁰ Some readers have had a different intuition about this. Relying on the thought that it is insulting not to trust somebody who is trustworthy, they conclude that trustworthiness *is* a favoring concept; that to be trustworthy is to deserve to be trusted. I think, rather, that the insult lies in withholding trust from a trustworthy agent in circumstances in which you have independent reason to repose trust in them if you deemed them trustworthy.

²¹ This is not to prejudge whether B’s reasons to engage in blaming activity could be grounded, in whole or part, in the facts that ground the fact of A’s blameworthiness.

recommend it. And other theoretical criteria *do* recommend it, including the hope and expectation that our existing responsibility concepts and associated nomenclature will fit tolerably well the normative functions that our responsibility concepts must serve. Therefore, I propose to pair our earlier analytic answer with respect to desert with the following analysis of blameworthiness:

The analytic answer (blameworthiness): What it is for an agent, A, to be blameworthy for an act, φ , is for it to be the case that A is rendered liable to negative or harsh treatment—“blaming practices”—to which A would otherwise be immune.

Assuming now that desert is a favoring concept and blameworthiness is a liability concept, resolution of the moral outcome luck puzzle turns upon the answers to Rosen’s grounding question as applied to each.

The grounding question (regarding outcomes and desert): Are the fortuitously realized or unrealized results of an agent’s wrongdoing among the determinants of the severity of the treatment or response that is favored for the agent (i.e., that possesses final value, or that somebody else ought to bring about) in virtue of their wrongdoing?

The grounding question (regarding outcomes and blameworthiness): Are the fortuitously realized or unrealized results of an agent’s wrongdoing among the determinants of the blaming practices to which the agent is liable in virtue of their wrongdoing?

The next two sections tackle these questions in order.

4. Negative Desert and Luck

What are the determinants of the character and severity of the punitive response that is favored (bears positive value, or ought to be inflicted) in virtue of an agent’s wrongdoing? This is the grounding question, as applied to negative desert. We can’t answer it straightaway. The explanatory question (or a somewhat more expansive variant) must be addressed first.

Whether or not H deserves more punishment (or other type of blame) than M does, the inquiry presupposes that both H and M deserve *some* punishment (or blame). Before we can intelligibly assess whether H deserves *more*, we need some grasp of why they both deserve something bad or disagreeable in the first place. Why does the fact that they tried to kill another human being without justification render it good that they

should suffer (or face deprivations, or shoulder burdens, etc.), or provide non-instrumental reason for others to inflict suffering or impose deprivations or burdens upon them? This is a notoriously hard question. Until we have a good handle on what makes negative desert plausible or defensible, or what its “moral underpinnings” are (Sher 1987, x), any attempt to identify the full determinants of the negative desert object (roughly, the deserved punishment) is flying blind.

Unfortunately, we lack a shared account of what makes desert sensible or warranting our allegiance. Worse, we lack even a well-developed menu of options to choose from. This is true even of positive desert objects, such as rewards. It is true in spades of negative desert objects, especially given determinist-inspired objections. How can the imposition of disagreeable consequences on wrongdoers be non-instrumentally good or right if the wrongdoers lacked freedom to do otherwise?²² Many people who do accept negative desert view it as a moral primitive, accessible to moral intuition or the product of fundamental moral commitments or convictions. I can’t rule that out, but it’s plainly “a theory of last resort.” (Rosen 2015, 71; cf. Moore 1997, 223). For one thing, if desert lies at unexplainable moral bedrock then it’s hard to see how we could resolve the grounding question except by consulting the same moral intuitions that reveal the truth about desert itself. Reasoned argument would seem fruitless.

I cannot proceed far in developing and defending a full account of desert in this article. Instead, I’ll sketch the rudiments of an answer to the explanatory question (posed of negative desert as a moral institution) that I find most promising, what I’ll call the “agency-maintenance” account of negative desert. I’ll then hazard some thoughts about what that answer, skeletal though it is, implies for the grounding question.

4.1. The point of desert

The agency-maintenance account I favor explains desert as playing a critical constitutive role in the formation and maintenance of selfhood. It draws upon a subtle and incisive account pressed by George Sher (1987, ch. 9). I see it as consisting of three core elements.²³

²² Derk Pereboom (2015, 281) seems to have a favoring conception of desert firmly in mind when observing, rightly in my judgment, that, of all forms or senses of moral responsibility, desert “is the sense most clearly threatened by” causal determinism.

²³ Although influenced by Sher’s account, mine departs from his principally in emphasizing the importance of free will, a (supposed) attribute of the self that is (curiously) absent from Sher’s analysis. It seems to me both that free will is a central aspect of the conception of the self at the heart of our moral system, and that that aspect is essential to make sense of, and to vindicate, “negative” or “retributive” desert. By marginalizing or

First, the moral system presupposes a certain conception of the moral self, one that, among other qualities, possesses free will, and continues over time. Whether, or in what respects, these presuppositions are true or false, the system's intelligibility and effective functioning depend upon their preservation. Second, desert's function, and thus its justification, is to assist in the construction, nourishing, and shaping of the sense of self upon which our moral system depends. The value of desert is thus both extrinsic and final. Third, desert carries out this function by tethering present and future selves to each other via the choices of the present self. It gives future selves strong stakes in the moral character of the willings of the present self by directing that consequences should obtain for those future selves. And by insisting that the obtaining of those results is just—indeed, that justice enjoins us to bring them about—desert proclaims agential freedom: Given the widely held premise that if people lack free will then they cannot possess negative desert, morality's embrace of negative desert affirms free will by *modus tollens*.

In sum, desert's value is grounded (at least in part) in the way it gives agents a stake in the exercises of their agency, thereby both helping to make agency function effectively in the construction of selfhood and strengthening dispositions to abide by morality's dictates. This is not a defense of desert from a perspective external to our system of moral reasoning and practice. It is the seeds of an account of the critical function that desert serves within that system.

4.2. Agency-maintenance and results

Assuming that something along the foregoing lines is correct—that desert is intelligible and attractive largely because of the role it plays in securing and strengthening the sense of selfhood that morality presupposes and depends upon—what are the implications for the grounding question? This too is disputable.

If an account that melds selfhood, agency, and desert is broadly on target, then the ground of negative desert would encompass those aspects of what a person has done that are fairly attributed to their agency by individual psychology and shared cultural practices that are defensible on reflection. Accordingly, if we could maintain a substantial and coherent sense of ourselves stripped down to that which we control, we might be

ignoring free will, Sher is driven (ch.5) to embrace Herbert Morris's (1968) "fair play" justification for retributive punishment, a celebrated analysis that has attracted fewer adherents over time. For a review of some criticisms, albeit in a rare recent contribution that leans supportive, see Westen (2016).

able to reap the benefits that the institution of desert makes possible without including factors that lie beyond our control in our desert base. We might be able to get by with a picture of the self constituted solely (or nearly so) by its willings alone.

But the antecedent of that last conditional is highly doubtful. It's not only, as Wolf (2001, 13) emphasizes, that *virtuous* agents count (many of) the results of their willings on their moral ledgers, as part of who they are and what they've done. More probably, most psychologically healthy individuals, virtuous or not, do so too. As Nagel observed (1979, 36-37), "the self which acts and is the object of moral judgment is threatened with dissolution by the absorption of its acts and impulses into the class of events. . . . [S]omething in the idea of agency is incompatible with actions being events or people being things." This being so, as I believe it to be, then if desert links and integrates temporal slices of a self by putting later selves "on the hook" for earlier selves' actions, the "actions" that matter must be broader than bare willings, or willings plus bodily movements, and extend to at least some of our willed movement's causal effects in the world.

Agreeing that "at least one connection between free acts and their consequences is internal to the notion of free agency itself," Sher (1987, 39-40) argues that "[i]f someone did *not* have to live with the predictable consequences of his choices—if he were able to go through the motions of deciding what to do, but was invariably shielded from his action's easily foreseeable results—then he would have only a semblance of freedom." That could be so, though I would put matters just a little differently: if an agent were shielded from their action's foreseeable and foreseen results, they would have only a semblance of selfhood. Desert helps preserve more robust selfhood by embracing results of our willings among the determinants of the treatments we deserve.

But only some of the results, not all of them, and it is a virtue of the agency-maintenance account that it holds promise of explaining why and which. Because this account of desert's moral underpinnings depends on the function it serves for members of a moral community, it turns on how we understand ourselves and our actions, and not (only) on timeless truths about, say, the nature of agency. It follows that whether results count among the desert base of deserved punishment will be and must remain significantly sensitive to folk wisdom.

The account thus makes sense of the fact that common law doctrines of causal responsibility are largely social, not metaphysical. Examples include the role of foreseeability and non-aberrant causal pathways in proximate causation, and the understanding that intentional intervening action by third parties "breaks the causal chain" even when foreseeable. The account

also allows for the possibility that whether results count in the negative desert base can vary depending on the wrongdoer's mental state (intention, foresight, recklessness, or negligence) with respect to the bad results that do or do not materialize, for the simple (if not sole) reason that these diverse mental states constitute or correspond to differences in the extent to which the agent has associated their self with the results, as by desiring them. Contrast H and M, the shooters who intend the death of V, with two other actors, H_N and M_N, who act negligently with respect to V_N's death. (The standard hypothetical involves careless driving: H_N and M_N drive identically in respect of the facts that constitute negligence, but H_N has the bad luck to hit and kill a pedestrian while M_N has the good luck not to.) The fact that we naturally describe H himself, and not only H's actions, by terms ("killer," "murderer") that make necessary reference to the results of that action, and that such agential descriptions are definitionally inapplicable to M, plausibly goes some distance toward explaining why the happenstance whether V was hit and killed counts among the determinants of H's and M's negative desert. In contrast, we probably resist describing H_N in terms (e.g., "killer," "manslaughterer") that refer to the results of H_N's actions, choosing instead to describe both H_N and M_N as negligent drivers, while also noting that H_N is a negligent driver who caused somebody's death. On the agency-maintenance account, these facts about ordinary thought and talk have both evidential and constitutive significance in determining the contours of outcome luck for negative desert.

* * *

This section's discussion is admittedly tentative. It is reasonably contestable whether (or which) results are determinants even assuming that the Sher-inspired selfhood-securing account of desert at which I gesture is correct. And the merits of that account are equally contestable. For both these reasons, I do not offer a confident bottom-line judgment about whether results count among the determinants of an agent's desert object. But it cannot be this section's task to develop and defend a complete vindicating explanation for desert as a moral concept with normative force. Its more limited ambition is to offer a promising sketch of such an explanation and thereby to show how resolving whether outcomes of an agent's action bear on what they deserve depends, first, on settling on an answer to the explanatory question as applied to negative desert as a functional concept within our moral system.

5. Blameworthiness and Luck

What conditions must obtain to render an actor morally liable to be subjected to blaming practices? Here's a common idea (e.g., Brink and Nelkin, 289). First, the actor must be a "moral agent," someone who is, in Gary Watson's (1996) sense, morally "accountable." Second, even an accountable agent is not liable to blame unless they were "at fault." Now, precisely what fault consists in is controversial—failure to satisfy a fair standard of conduct fairly enforced, conscious violation of a moral norm, and so forth. But that an agent isn't blameworthy—isn't morally liable to being blamed—unless at fault in one or another sense is common ground.

Also common to all the varied conceptions of the fault that grounds blameworthiness is that it involves matters "internal" to the agent. Subject to a few possible clarifications or qualifications,²⁴ it turns on features of the agent's will or the way that they deliberated or failed to deliberate. Because blameworthiness is a scalar not binary concept, it might seem to follow that the extent of one's blameworthiness must be entirely determined by the extent of one's fault. It might seem to follow, in other words, that the character and severity of the blaming practices to which one is liable when blameworthy are entirely determined by the internal factors that constitute fault, whatever they may be. And if this is so, then the results of an agent's conduct are categorically irrelevant to their blameworthiness, which is to say that blameworthiness must be entirely insensitive to luck in the way things turn out. In our motivating hypothetical, H and M would be equally blameworthy. (And same too for H_N and M_N.) Wolf's "rationalist position" (2001, 6-7) captures this familiar line of argument.

The conclusion is too quick, however, for the fact (if true) that fault is the sole necessary condition for an agent to be blameworthy does not entail that it is the sole determinant of the scope or extent of their blameworthiness. Although this point might be obvious enough once asserted, its force becomes more apparent when we note an ambiguity built into the notion of degrees of blameworthiness, and thus disambiguate two respects in which blameworthiness is scalar.²⁵ Take two agents both of

²⁴ I am especially thinking that external standards of proper behavior may bear constitutively on the fault that constitutes blameworthiness in cases of (inadvertent) negligence, *if* blameworthiness can consist in negligence, itself a much-debated matter. See Berman (forthcoming 2022a).

²⁵ The distinction is hinted at in Jensen (1993) and keenly developed in Enoch and Marmor (2007, 412-17). My analysis on this point is in broad accord with theirs, although I

whom are blameworthy for some (faulty) conduct. One could be more or less blameworthy than the other in terms either of (a) the magnitude or severity of blame to which they are liable, or (b) the extent of their liability to blame. So there are two possible grounding questions about blameworthiness, not one: (a) what determines how much blame the actor is liable to? And (b) what determines how much the actor is liable to blame? These are different questions that could provoke different answers.

The remainder of this section explicates and defends these two dimensions of the scalarity of blameworthiness, draws out their straightforward implications for moral outcome luck, and addresses one concern the picture might provoke. What this section does not do is spend any time pondering what need we have for a concept (blameworthiness) that serves a liability function—an omission that would be unworthy of note but for Section 4's insistence that we do need to investigate the normative grounding or justification for a concept (desert) that serves a favoring function. The explanation for this asymmetry is simple. As Section 4 emphasizes, the intelligibility or justifiability of negative desert is nonobvious and reasonably contested. In contrast, the intelligibility and justifiability of blameworthiness is obvious and noncontroversial. Given what we know about human psychology and social dynamics, blaming practices that are disagreeable to the blaming patient cannot be wholly, or even substantially, eliminated. Accordingly, a morality sensitive to the human condition must recognize the permissibility of at least some such practices, in which case we have plain need for a moral operator that performs a liability function.

5.1. Two respects in which blameworthiness is scalar

In our actual penal practices and everyday lives, we recognize very many factors as bearing on the extent, duration, and severity of appropriate or permissible blaming practices. Collectively, we consider: the gravity of the wrong done or the importance of the norm violated; the extent of the agent's departure from the standard of conduct in cases where departures can be measured in degrees; whether the agent knew their conduct was wrong; whether the agent had done it before; the agent's background and

do not share their contention (413) that “[q]uestions about blameworthiness are about the truth or falsehood of attributions of blame.” It seems to me that what we attribute to an agent when engaging in “blame-related reactions” is not “blame” but “blameworthiness,” yet we cannot substitute the latter for the former on pain of circularity. Furthermore, Enoch and Marmor overlook what I am claiming is a critical distinction between desert and blameworthiness and therefore do not consider the possibility that the denial of free will threatens the former but not the latter (435).

upbringing and the breadth of their opportunities to avoid wrongdoing; the treatment accorded other, similarly situated faulty actors; the “standing” of those who would impose blame; the likelihood that a severe response would have a positive effect on future behavior by this agent or by similarly situated others; and other factors too. Unless our accepted practices are in truly bad order (admittedly possible), it must be that the severity of blame to which we make ourselves liable by our fault is not fully determined by the internal factors that determine fault, thereby licensing blame.

What are the alternatives? Conceivably, the relevant factors determine permissible blame via a complex formula or function. But the history of efforts to distill such a function does not encourage optimism. Much more plausible is that an agent’s faulty behavior renders them liable to “reasonable” blaming practices, where reasonableness, here as elsewhere, is a function of a plurality of factors and resists significant precisification.

Even if so, suppose now that the blameworthy actor, albeit a moral agent and not somebody exempt from the reactive attitudes, suffers from significant cognitive or volitional impairments. For example, the agent: suffers from mental disease or disorder that interferes with normative reasoning or behavioral regulation, is an adolescent, has a very low I.Q., acts in a state of reason-clouding anger (provoked by adequate provocation), intense fear, or extreme sleep deprivation, or is (involuntarily) intoxicated. Most commentators believe that all or some of such factors mitigate the agent’s blameworthiness; some are subjects of full or partial excuse in the criminal law. But these factors do not operate—certainly not exclusively, and I think not primarily—as considerations that bear on what blaming practices would be all-things-considered reasonable, as by weighing toward leniency. They operate by reducing the agent’s liability to blaming and punitive responses. This is largely what is at stake when commentators characterize persons who exhibit these defects or conditions as “less responsible.”

Combining these two dimensions of variability produces this first sketch of the fuller picture of blameworthiness: To be blameworthy is to be liable to blame in virtue of one’s being at fault. The character and severity of blame to which the faulty actor renders themselves liable is determined by what is reasonable given the agent’s conduct, the consequences of that conduct, and the relevant circumstances, discounted by any diminution in the agent’s liability to blame grounded, paradigmatically, in defects in the agent’s settled or episodic rationality or systems of intrapersonal integration or control. Many mechanical systems incorporate two devices to adjust output (e.g., volume, heat, power): one that permits gross adjustment, a second for fine-tuning. In roughly analogous fashion, the

magnitude of hostility or disagreeability to which one is liable in virtue of one's faulty conduct is determined by all the factors that bear on the reasonableness of a response, discounted by the degree of the actor's liability to such responses.

A vignette illustrates. Suppose that Ringleader, Schnook, and Juvenile together commit Crime. It was Ringleader's idea, of course. Schnook joined as a most reluctant and regretful second fiddle, and only after long resisting Ringleader's wheedling and importuning. Juvenile signed on willingly, even enthusiastically, but is an adolescent. When arrested, Schnook and Juvenile argue strenuously that they are "less blameworthy" than Ringleader, and should be punished less severely—Schnook because he was less "culpable," in a broad sense bespeaking a less "vicious will," and Juvenile because he was less emotionally and psychologically mature. I assume that readers can fill in details (concerning the age or maturity of the protagonists, Ringleader and Schnook's interactions and the influence the one had over the other, the Crime itself, and so on) in ways that would make Schnook's and Juvenile's arguments compelling. But while both claims are couched in terms of the claimant's lesser blameworthiness, I anticipate that many readers will share the sense that these claims have distinctly different texture. The account I've just sketched aims to reflect that difference: Schnook maintains that they're liable to less blame than is Ringleader; Juvenile maintains that they're less liable to blame than is Ringleader.

Is the proposed distinction between how much one is liable to blame and how much blame one is liable to sterile? Does anything turn on the fact that Juvenile's claim is one of the latter and not (only) of the former? Would it matter if you became persuaded that Schnook's claim also concerned liability? There are reasons to believe that it does matter. The gist of the case for Juvenile is not only or chiefly that it's unreasonable to disregard the difference in their maturity, but that we are morally required to accommodate it, that Juvenile has a valid claim to diminution or attenuation of his liability to blaming treatments. Plausibly, then, punishing Ringleader, Schnook, and Juvenile equally would treat both Schnook and Juvenile unreasonably hence unfairly, but would violate the moral rights only of Juvenile. Or, shifting attention away from state-inflicted punishment and toward social practices, maybe Victim of Crime, or others in the community, may be entitled, when calibrating their blaming responses, to disregard differences between Ringleader and Schnook, but not between Ringleader and Juvenile.

5.2. The relevance of results

This two-part account of the ways in which blameworthiness comes in degrees neatly explains our divergent intuitions about whether results of an agent's conduct affect their blameworthiness.

Question (a): Are the fortuitously realized or unrealized results of an agent's faulty or wrongful conduct among the determinants of the blaming practices to which the agent is liable in virtue of their fault? Surely yes. If being at fault makes one liable to "reasonable" blaming practices, then, whatever the full set of considerations that bear on the propriety of a blaming response might consist in, it's hard to dispute that at least some results of the faulty agent's conduct—what the agent does and causes—count within that set.²⁶ Experiencing resentment and indignation, and engaging in practices that express those sentiments and allow for the channeled discharge of retaliatory impulses, are all part of the human condition. But it is pathological to harbor as much resentment, or to indulge in equally severe chastisement or retaliation, regardless of whether another's faulty behavior caused you or those you care about any harm at all. So if H and M are equally liable to reasonable blame, H is nonetheless liable to more severe blame or retaliation, including blame that is more severe in virtue of extending for greater duration, just because the amount of reasonable outrage and hostility provoked by H's killing of V is greater than that provoked by M's failed attempt to kill V. Furthermore, if results *are* part of the desert base (the question left open in Section 4), and if reasonable blaming practices include the infliction of whatever censure or hard treatment may be deserved, then, again, the severity of blame to which a faulty agent is liable is greater if the results of the faulty conduct are worse.

²⁶ This contention is largely supported by findings from the moral cognition literature, though I don't place much weight on it. After all, given that I am arguing that the moral-conceptual landscape harbors more distinctions than the relevant philosophical literature has yet recognized, it's almost inevitable that I'd find the empiricists' survey instruments insufficiently nuanced. For example, one sophisticated recent contribution that fairly criticizes previous studies for invoking "a staggering variety in types of judgments," asks respondents "how blameworthy" contrasting agents are, how "permissible" their actions were, and "how much punishment" each "deserves." Kneer and Machery (2019, 333, 334); see also Cushman (2008, 358). My analysis suggests that respondents highly attuned to philosophically relevant distinctions might give different answers to the questions "how much punishment does the agent *deserve*?" "how much punishment would be *reasonable* to impose?" and "how much punishment would be *permissible* to impose?" To illustrate, a sensible respondent might believe that the reasonable punishment for Schnook and Juvenile is roughly the same but that the permissible punishment for Schnook is greater. And if such a respondent is a negative desert skeptic, they might also believe that none of the actors deserves any punishment at all.

This is not to say that the reactive treatment that the agent deserves constitutes the full measure of what is reasonable. Even assuming that reasonable blaming responses are responses we have reason to undertake, those reasons are not nearly limited to that of effectuating deserved treatment for its own sake; they include deterrent, expressive, and solidaristic reasons, among others.²⁷

Moreover, because the agent subjected to blame could have avoided the influence of outcome luck by not acting in a faulty manner to start, the fact that the degree of blame to which they are liable is sensitive to some facts beyond their control does not seem unfair: the agent has assumed that risk. Invoking Ronald Dworkin's (1981) distinction between "brute luck" and "option luck," Michael Otsuka (2009, 375) puts the point well: "A plausible requirement of fairness would, at most, rule out unavoidable blameworthiness. It would not also rule out cases of moral luck in which one's degree of blameworthiness is a function of factors beyond one's control but where one could have avoided being blameworthy altogether."²⁸

Question (b): Are the fortuitously realized or unrealized results of an agent's wrongdoing among the determinants of the agent's liability to blaming practices? Plainly not. Although what determines an agent's responsibility in the accountability or liability-discounting sense is far from settled, the standard accounts focus squarely on defects in agency, such as deficits in reason-responsiveness or executive control. If it is stipulated that the only difference between H and M concerns the downstream, external consequences of their pulling of their respective gun triggers, then they are equally "accountable," equally liable to blame.

Moore (1997, 212) interprets Nagel thusly: "The problem of moral luck . . . is how we can justify holding people more responsible for causing harm than for merely intending [harm]." Distinguishing two scalar notions baked into the concept of blameworthiness permits a clean solution. The question should not be whether an agent who causes harm is "more responsible" than one who, all else equal, tries to cause harm but fails. Certainly, they are equally responsible in the sense, among others, of being equally liable to blaming responses. The question should be whether we can justify inflicting more severe blame on those who cause harm than on those who don't, despite their being equally liable to blaming treatment. Certainly, we can. Although H and M are equally responsible, hence

²⁷ Accord Enoch and Marmor (2007, 413). I am grateful to an anonymous referee for encouraging this clarification.

²⁸ The same insight underlies David Lewis's (1989) proposed penal lottery.

equally liable to reasonable blame, the amount and character of blame that would be reasonable to subject each to is not the same.

5.3. A loose end: reasonable blame and principles of proportionality

The proposal that to be liable to blame in virtue of one's fault is to make oneself fairly liable to "reasonable" blaming practices whose character and severity are determined by more than fault alone provokes a natural worry. Suppose a case involving very little fault but immensely bad consequences. If the public's anger has been greatly aroused, and if the welfarist benefits of deterring similar future acts are also great, then the imposition of severe blame, including severe punishment, will turn out to be all-things-considered reasonable at least sometimes. Call this upshot "the objectionable conclusion." If my parsing of the concepts yields this conclusion, that might seem a blot on the account.

To be sure, the objectionable conclusion *is* objectionable: assumption of risk has its limits as well as its force. To guard against it, nearly all philosophers of punishment agree that punishment is constrained by a "principle of proportionality," differing chiefly on that principle's correct content—for example, that punishment must not be disproportionate to the offender's "guilt," (Mackie 1982, 4) or the "gravity of the offense" (von Hirsch 1992, 55), or can be no greater than the minimum necessary to preserve civil order (Deigh 2018, 239). One family of principles that have attracted broad support maintains that punishment should not be disproportionately severe relative to the agent's "culpability" (alone, or in combination with "wrong" or "harm"), where culpability is a measure of the agent's ill will, or their disregard of interests and values that demand respect. (Culpability is a kind of fault, but only one kind of fault. There is no faultless culpability, but plenty of nonculpable faultiness.) I am sympathetic to a principle along these lines.

This is not the place to elaborate or defend a principle with any particular content; I take a stab at it in Berman (2022b). The important point for now is only that debates over principles of penal proportionality, or over other side constraints on the just infliction of state punishment, are not best conceived as debates about the contours of our moral concepts. Rather, they reflect substantive disagreements of political morality, first-order claims about the ethics or justice of state action. It is no embarrassment to a conceptual map, this or any other, that it does not exclude all morally disagreeable practices and states of affairs on purely conceptual grounds, and that we must call on substantive moral principles to perform that trick.

Conclusion

This article investigates how, if at all, moral outcome luck affects or determines an agent's blameworthiness and negative desert. In addressing that question, it identifies a premise that appears widely presupposed in the moral luck debate: that the determinants of blameworthiness and desert are the same. I call that premise the equivalence thesis, and argue that it is false. In rejecting that premise, this article sheds light on the tantalizing possibility that results, and thus luck that attends their realization or non-realization, can matter to one concept but not the other. Nonetheless, the paper ends up resisting this neat and clean resolution to the puzzle of outcome luck. It maintains that results probably bear both on the magnitude or severity of the blame to which the agent is liable when blameworthy and on the magnitude or severity of the adverse treatment that the agent deserves, though not on the extent of the agent's liability to blaming practices.

Because those last conclusions are modestly more equivocal than reader and author alike might wish, it bears emphasis that, while this article addresses the purported equivalence of blameworthiness and desert as they concern the relevance of outcomes, the equivalence thesis itself is not of such limited import. That thesis concerns the more general question of whether the determinants and measure of an agent's blameworthiness for an act are the same as the determinants and measure of an actor's desert for the same act. The negative answer I defend to that question has implications for moral outcome luck, but not only for that. The truth or falsity of the equivalence thesis is likely to matter anytime an agent's desert or blameworthiness bears on the permissibility of criminal punishment. A clear distinction between the functions that desert and blameworthiness serve, and a clear understanding of their respective determinants, might shed light on debates: over the nature and bounds of responsibility-based side constraints on the infliction of punishment; over the permissibility of punishing people for negligence; over the proper scope of a defense of moral ignorance; and over much else besides. Although this one paper does not examine those other topics, it is worth keeping in mind that its initial and most important claims—desert serves a favoring function, blameworthiness serves a liability function, and the determinants of these two concepts are non-identical—will bear implications for a very broad range of topics within the philosophy of blame and punishment.

Works Cited

- Arpaly, Nomy. 2003. *Unprincipled Virtue: An Inquiry into Moral Agency*. Oxford: Oxford University Press.
- Berman, M.N. (2022). Negligence and Culpability: Reflections on Alexander and Ferzan. *Criminal Law and Philosophy*. Advanced online publication. doi: 10.1007/s11572-021-09583-8.
- Berman, M.N. (2021). Proportionality, Constraint, and Culpability. *Criminal Law and Philosophy*, 15, 373-91. doi: 10.1007/s11572-021-09589-2.
- Berman, M.N. (2016). Modest Retributivism. In K.K. Ferzan & S.J. Morse (Eds.), *Legal, Moral, and Metaphysical Truths: The Philosophy of Michael S. Moore* (pp. 35-47). Oxford: Oxford University Press.
- Berman, M.N. (2013). Rehabilitating Retributivism. *Law and Philosophy*, 32, 83-108. doi: 10.1007/s10982-012-9146-1.
- Berman, M.N. (2011). Two Kinds of Retributivism. In R.A. Duff & S. Green (Eds.), *The Philosophical Foundations of Criminal Law* (pp. 433-57). Oxford: Oxford University Press.
- Brink, David O., and Dana K. Nelkin. 2013. "Fairness and the Architecture of Responsibility," in *Oxford Studies in Agency and Responsibility*, edited by David Shoemaker. Oxford: Oxford University Press: 284-313.
- Coates, D. Justin, and Neal A. Tognazzini. 2013. "The Contours of Blame," in *Blame: Its Nature and Norms*, edited by Justin D. Coates, and Neal A. Tognazzini. Oxford: Oxford University Press: 3-26.
- Christopher, Russell. 2004. "Does Attempted Murder Deserve Greater Punishment than Murder? Moral Luck and the Duty to Prevent Harm." *Notre Dame Journal of Ethics and Public Policy* 18: 419-35.
- Cupit, Geoffrey. 1996. *Justice as Fittingness*. Oxford: Oxford University Press.
- Cushman, Fiery. 2008. "Crime and punishment: Distinguishing the roles of causal and intentional analyses in moral judgment." *Cognition* 108: 353-80.
- Deigh, John. 2018. *From Psychology to Morality: Essays in Ethical Naturalism*. Oxford: Oxford University Press.
- Domskey, Darren. 2004. "There is No Door: Finally Solving the Problem of Moral Luck." *Journal of Philosophy* 101: 445-464.

- Dworkin, Ronald. 1981. "What is Equality? Part 2: Equality of Resources." *Philosophy and Public Affairs* 10: 283-345.
- Edwards, James, and Andrew Simester. 2019. "Crime, Blameworthiness, and Outcomes." *Oxford Journal of Legal Studies* 39: 50-73.
- Enoch, David, and Andrei Marmor. 2007. "The Case Against Moral Luck." *Law and Philosophy* 26: 405-36.
- Feldman, Fred. 1997. *Utilitarianism, Hedonism, and Desert*. Cambridge: Cambridge University Press.
- Feinberg, Joel. 1970. *Doing and Deserving*. Princeton, N.J.: Princeton University Press.
- Freiman, Christopher, and Shaun Nichols. 2011. "Is Desert in the Details?" *Philosophy and Phenomenological Research* 82: 121-33.
- Graham, Peter A. 2014. "A Sketch of a Theory of Moral Blameworthiness." *Philosophy and Phenomenological Research* 88: 388-409.
- Greco, John. 1995. "A Second Paradox Concerning Responsibility and Luck." *Metaphilosophy* 26: 81-96.
- Greenberg, Mark. 2014. "The Moral Impact Theory of Law." *Yale Law Journal* 123: 1288-1342.
- Hartman, Robert J. 2019. "Moral Luck and the Unfairness of Morality." *Philosophical Studies* 176: 3179-3197.
- Hurka, Thomas. 2001. "The Common Structure of Virtue and Desert." *Ethics* 112: 6-31.
- Jensen, Henning. 1993. "Morality and Luck," in *Moral Luck*, edited by Daniel Statman. Albany: SUNY Press: 131-40.
- Kadish, Sanford. 1994. "Forward: The Criminal Law and the Luck of the Draw." *Journal of Criminal Law and Criminology* 84: 2183-2237.
- Kagan, Shelly. 2012. *The Geometry of Desert*. Oxford: Oxford University Press.
- Kershnar, Stephen. 2010. *Desert and Value*. Lowman, MD: Rowman & Littlefield.
- Khoury, Andrew C. 2018. "Criminal Attempts and the Penal Lottery." *Australasian Journal of Philosophy* 96: 779-792.
- Kneer, Markus, and Edouard Machery. 2019. "No luck for moral luck." *Cognition* 182: 331-48.

- Lee, Youngjae. 2009. "Recidivism as Omission: A Relational Account." *Texas Law Review* 87: 571-622.
- Levy, Ken. 2005. "The Solution to the Problem of Outcome Luck: Why Harm is Just as Punishable as the Wrongful Action that Causes it." *Law & Philosophy* 24: 263-303.
- Lewis, David. 1989. "The Punishment that Leaves Something to Chance." *Philosophy and Public Affairs* 18: 53-67.
- Mackie, J.L. 1982. "Morality and the Retributive Emotions." *Criminal Justice Ethics* 1: 3-10.
- McKenna, Michael. 2013. "Directed Blame and Conversation," in *Blame: Its Nature and Norms*, edited by Justin D. Coates, and Neal A. Tognazzini. Oxford: Oxford University Press: 119-40.
- Moore, Michael S. 1997. *Placing Blame*. Oxford: Oxford University Press.
- Morris, Herbert. 1968. "Persons and Punishment." *Monist* 52: 475-501.
- Nagel, Thomas. 1979. *Mortal Questions*. Cambridge: Cambridge University Press.
- Nelkin, Dana. 2019. "Moral Luck," in *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, <https://plato.stanford.edu/entries/moral-luck/#KinMorAss>.
- Olsaretti, Serena. 2003. "Introduction: Debating Desert and Justice," in *Desert and Justice*, edited by Serena Olsaretti. Oxford: Oxford University Press: 1-24.
- Otsuka, Michael. 2009. "Moral Luck: Optional, Not Brute." *Philosophical Perspectives* 23, *Ethics*: 373-388.
- Pereboom, Derk. 2015. "A Notion of Moral Responsibility Immune to the Threat from Causal Determinism," in *The Nature of Moral Responsibility: New Essays*, edited by Randolph Clarke, Michael McKenna, and Angela M. Smith. Oxford: Oxford University Press: 281-96.
- Richards, Norvin. 1986. "Luck and Desert." *Mind* 65: 198-209.
- Rosebury, Brian. 1995. "Moral Responsibility and 'Moral Luck.'" *Philosophical Review* 104: 499-524.
- Rosen, Gideon. 2015. "The Alethic Conception of Moral Responsibility," in *The Nature of Moral Responsibility: New Essays*, edited by Randolph Clarke, Michael McKenna, and Angela M. Smith. Oxford: Oxford University Press: 65-87.

- Ross, W.D. 1930. *The Right and the Good*. Oxford: The Clarendon Press.
- Sher, George. 1987. *Desert*. Princeton: Princeton University Press.
- Sverdlik, Steven. 1988. "Crime and Moral Luck." *American Philosophical Quarterly* 25: 79-86.
- Tadros, Victor. 2011. *The Ends of Harm: The Moral Foundations of Criminal Law*. Oxford: Oxford University Press.
- Tadros, Victor. 2017. "Moving Mountains: Variations on a Theme by Shelly Kagan." *Criminal Law and Philosophy* 11: 393-405.
- Thomson, Judith Jarvis. 1989. "Morality and Bad Luck." *Metaphilosophy* 20: 203-221.
- von Hirsch, Andrew. 1992. "Proportionality in the Philosophy of Punishment." *Crime and Justice: A Review of Research* 16: 55-98.
- Watson, Gary. 1996. "Two Faces of Responsibility." *Philosophical Topics* 24(2): 227-248.
- Westen, Peter. 2016. "Retributive Desert as Fair Play," in *Legal, Moral, and Metaphysical Truths: The Philosophy of Michael S. Moore*, edited by Kimberly Kessler Ferzan, and Stephen J. Morse. Oxford: Oxford University Press: 63-78.
- Wolf, Susan. 2001. "The Moral of Moral Luck." *Philosophic Exchange* 31: 4-19.
- Zimmerman, Michael. 2015. "Varieties of Moral Responsibility," in *The Nature of Moral Responsibility: New Essays*, edited by Randolph Clarke, Michael McKenna, and Angela M. Smith. Oxford: Oxford University Press: 45-64.
- Zimmerman, Michael. 2002. "Taking Luck Seriously." *Journal of Philosophy* 99: 553-76.