

University of Michigan Law School

University of Michigan Law School Scholarship Repository

Reviews

Faculty Scholarship

2020

Demystifying Desert

Gabriel Mendlow
mendlow@umich.edu

Available at: <https://repository.law.umich.edu/reviews/150>

Follow this and additional works at: <https://repository.law.umich.edu/reviews>



Part of the [Criminal Law Commons](#), and the [Law and Philosophy Commons](#)

Recommended Citation

Mendlow, Gabriel S. "Demystifying Desert." *The Journal of Ethics* 24 (2020): 287-294.

This Review is brought to you for free and open access by the Faculty Scholarship at University of Michigan Law School Scholarship Repository. It has been accepted for inclusion in Reviews by an authorized administrator of University of Michigan Law School Scholarship Repository. For more information, please contact mlaw.repository@umich.edu.



Demystifying Desert

Gabriel S. Mendlow¹ 

Received: 25 October 2019 / Accepted: 26 April 2020 / Published online: 9 May 2020
© Springer Nature B.V. 2020

Abstract

In his penetrating book on the criminal culpability of children, Gideon Yaffe advances a novel theory of desert. According to the theory, the punishment you deserve for committing a given crime is the punishment the prospect of which would have led you to deliberate correctly about how to act, had that punishment been presented to you beforehand as an inevitable consequence of your committing the crime. Although fascinating and ambitious, Yaffe’s theory of desert struggles as an account of who deserves what, and it falls short of explaining why and how desert is normatively significant—why and how desert matters.

Keywords Desert · Punishment · Culpability

1 Introduction

It is widely believed that the guilty deserve to suffer. But what kind and degree of suffering do the guilty deserve? And why does their deserving to suffer entail that the state or anyone else ought to *make* them suffer? I find these questions difficult—so difficult that I am skeptical of efforts to justify state punishment by reference to what criminals deserve, even as I am reluctant to deny that at least some criminals deserve to suffer some kind of hardship. Desert seems at once undeniable and obscure. Gideon Yaffe rightly calls it “mysterious” (2018: 98).

Two questions about the mystery of desert will be my focus in this essay. The first is what exactly the guilty deserve. It is a notoriously difficult question, familiar from sentencing theory as the problem of identifying the proper “anchoring” points for a sentencing scheme (von Hirsch 1993: 36). The second question is why and how desert is normatively significant—why and how desert matters. Why should people get what they deserve, and how urgently should they get it? Must they get what they deserve at all costs, or do the demands of desert sometimes give way to competing values and constraints? How weighty or stringent are the demands of desert? And

✉ Gabriel S. Mendlow
mendlow@umich.edu

¹ University of Michigan, 625 S. State St., Ann Arbor, MI 48104, USA

which agent or institution has the authority to satisfy them, the authority to give people their just deserts?

In his penetrating book on the criminal culpability of children, Yaffe (2018) offers a novel *corrective* theory of desert. It is a fascinating and ambitious theory, and it promises answers to some of the difficult questions posed above. Yaffe's principal claim is that

to give someone what they most deserve for past wrongdoing is to attach something to the [wrongful] act thanks to which the act has, for the agent, the very same reason-giving properties that it has, without that thing attached to it, to a better agent. If a year in prison is deserved for robbery, it is because a robbery with a one-year prison term attached to it has for the agent the same reason-giving properties as that same token robbery, with no such consequence attached to it, has for the good person who would not commit it.... [T]o give someone what he most deserves is, in a sense, to *correct* something.... If a good person takes a robbery to be unthinkable, for instance—he does not include it, in his deliberations, even among the acts he considers—then the robber gets what he deserves when the robbery, given a year in prison, would be, for him, unthinkable. Of course, “unthinkability” is just an example of a form of treatment of an act in deliberation. What is important is that the agent treats the act with the putatively deserved thing attached to it, in deliberation, in the same way that the good agent treats the act without that thing attached to it. (2018: 103)

According to this theory of desert, the punishment you deserve for committing a given crime is the punishment the prospect of which would have led you to deliberate correctly about how to act, had that punishment been presented to you beforehand as an inevitable consequence of your committing the crime. To deliberate correctly is to assess the reasonableness of committing the crime the way a morally “ideal agent” would assess its reasonableness absent a threat of punishment.¹

Yaffe's corrective theory of desert is potentially a significant philosophical advance. It provides what might seem a theoretically precise account of what wrongdoers deserve: they deserve exactly the sort of treatment that would have induced sound reasoning if it had been threatened as an unavoidable consequence of wrongdoing. The corrective theory also provides at least a partial explanation of desert's normative significance: the reason why people ought to get what they deserve is roughly that they ought to deliberate well, to deliberate the way an ideal agent would. I find these proposals intriguing, both for their specificity and for their ostensible explanatory power. But I wish to raise several doubts about their soundness.²

¹ “In the context of criminal law,” Yaffe explains, “the ideal agent is the agent who has modes of recognition and weighing of reasons thanks to which he grants as much reason-giving weight to the relevant features as is necessary for the act of refraining from [a given crime] to be more supported by reasons than the [crime itself is]” (2018: 108).

² In the passage quoted above, Yaffe advances a theory of *non-comparative* desert, a theory of what it is for a given punishment to be deserved outright, as opposed to deserved more or less than some other punish-

2 Who deserves what?

Whether Yaffe's corrective theory demystifies desert depends at least on how perspicuously we can articulate the notion of deliberative similarity. What does it mean for one agent to deliberate about an action "in the same way" that another agent does? As an example, Yaffe likens the way an ideal agent deliberates about an unpunishable robbery to the way a hypothetical defendant deliberates about a robbery coupled with a year in prison: each regards the prospective action as "unthinkable" (2018: 103). This comparison leaves the notion of deliberative similarity obscure, however. The two agents do not actually deliberate about the robbery in the same way, or even in roughly similar ways. The ideal agent treats the robbery as morally abhorrent, whereas the hypothetical defendant treats the robbery coupled with a year in prison as pointlessly self-destructive. These two forms of rational regard are not the same in any respects but valence and stringency. They are probably not even the same in stringency if stringency is counterfactually robust: the ideal agent presumably would not commit a robbery for any amount of money, while the hypothetical defendant might well commit a robbery

Footnote 2 (continued)

ment. When he formulates the theory more formally, the thesis he articulates is about *comparative* desert, about what it is for one punishment to be more deserved than another. The more formal thesis, which Yaffe calls "Desert as Isomorphism in the Space of Reasons," is this: "For C-ing, D deserves P1 more than P2 if and only if the form of rational support, relative to the ideal agent, enjoyed by C is more similar in relevant respects to the form of rational support, relative to D, enjoyed by C&P1 than that enjoyed, relative to D, by C&P2" (2018: 102). Put differently, D deserves P1 for C-ing more than D deserves P2 for C-ing if and only if, as between the two punishments, threatening D beforehand with the certainty of P1 would have done a better job of getting D to reason properly about whether to C than threatening D with the certainty of P2 would have done. Yaffe does not argue for a corresponding thesis about non-comparative desert, about what it is for a given punishment to be deserved outright. But the passage I quoted at the beginning makes clear that Yaffe envisions his corrective theory as encompassing comparative and non-comparative desert alike. Moreover, Yaffe elsewhere asserts a version of the corrective theory that is straightforwardly non-comparative (e.g., 2018: 153). The non-comparative version of the theory holds that D deserves P for C-ing if and only if threatening D with the certainty of P would have caused D to assess the reasonableness of C-ing the way an ideal agent assesses its reasonableness absent a threat of punishment.

The non-comparative thesis is my principal target in this essay. The non-comparative thesis speaks more directly to the question of what wrongdoers deserve, and it is probably sound *if the comparative thesis is*. Yaffe acknowledges the connection himself:

[T]o say that P1 is deserved, period, and not just more deserved than something else, is to say that the very same relation is to be found between P1 and C that we find more of when we find that P1 is more deserved for C than P2 is. This is just to say that the concept of non-comparative desert functions like any threshold concept. To be tall is to be taller than a person who is at a certain height threshold. For a form of treatment in response to wrongdoing to be deserved is for it to be more deserved than a form of treatment at a certain threshold. (2018: 102)

In other words, if P1's being more deserved than P2 is a matter of D's rational assessment of C&P1 being more similar to the ideal agent's rational assessment of C than D's rational assessment of C&P2 is, then it stands to reason that P's being deserved outright is a matter of D's rational assessment of C&P being *similar enough* to the ideal agent's rational assessment of C—the degree of similarity having exceeded some threshold. Specifying that threshold is a difficult task. It is a task Yaffe does not undertake (2018: 101), as his aim is to show that children deserve less punishment than adults do, not to show just how much (or how little) punishment children deserve outright.

at the cost of a year in prison if the haul were truly enormous. Yaffe's leading example thus leaves us without a clear conception of deliberative similarity.

Even if we rested content with a vague conception, we would still have reason to question whether the corrective theory offers a plausible account of desert. For the deliverances of the corrective theory run counter to commonsense intuitions about who deserves what. One especially discordant implication is that the penalty most criminals deserve for most crimes is extraordinarily mild. Why? Because most criminals would find most crimes "unthinkable" if they were sure to be caught and punished *even if the penalties were slight*. To my mind, Yaffe greatly underestimates the typical offender's aversion to certain punishment. Yaffe's hypothetical robber regards a robbery as "unthinkable" only when it is bound to result in an entire year in prison (2018: 103). I submit that most actual robbers would deem a given robbery unthinkable if it were coupled with anything but the mildest of punishments. Robbers rob because they think they'll get away with it. Greed motivates them to risk the *possibility* of punishment, not to suffer the certainty of punishment in exchange for the proceeds of a robbery. The certainty of any amount of punishment—even just a day in jail—is probably enough to dissuade all but the most hardened or hot-blooded of criminals from committing all but the most lucrative or viscerally satisfying of crimes. The corrective theory therefore implies that most criminals deserve no more than a day in jail for most crimes, an implication I find counterintuitive.

Less counterintuitive but still puzzling is a second concrete implication of Yaffe's theory: that a deserved punishment must be *aversive*, must be something the offender does not want. Only an aversive punishment can be truly deserved, implies Yaffe's theory, because only the prospect of aversive treatment can lead a would-be offender to deliberate about wrongful acts the way an ideal agent would. (Ideal agents presumably regard wrongful acts as highly unattractive—as "unthinkable" or something close to it.) This implication leaves the corrective theory with a puzzle I will not venture to unravel: what kind of sanction is deserved by conflicted and remorseful offenders who crave their just deserts, or by offenders who engage in acts of civil disobedience hoping to be caught and punished?

Yaffe seeks to preempt this sort of scrutiny by denying outright that our commonsense intuitions about who deserves what should carry any theoretical weight. Intuitions about desert are easily confused with intuitions about revenge, claims Yaffe, "[a]nd so there is really no way to know whether an account [of desert] that conforms to intuition is conforming to the right set" (2018: 101). That is a fair point, although it cuts against the overall thrust of Yaffe's argument. If our intuitions about desert really are as untrustworthy as Yaffe says, then we probably should not appeal to desert at all when seeking to justify legal punishment, a practice by which the state causes tremendous suffering, usually without immediate need.

3 Why and how does desert matter?

If the corrective theory struggles as an account of who deserves what, it does not for that reason fail as an account of desert's normative significance. A theory's success or failure in this regard depends most fundamentally on whether it

can help us understand why a person's deserving some treatment is a reason (of some still-to-be-determined weight and stringency) for the person to receive that treatment.

One theory that provides a clear answer to this question is the *unfair advantage* theory of retributive punishment, according to which a wrongdoer's punishment is deserved if it cancels the advantages that the wrongdoer obtained unfairly through her wrongdoing (see, e.g., Davis 1986). The theory has a well-known problem: some wrongdoing is entirely self-destructive, garnering no advantage that punishment can cancel (Yaffe 2018: 100). Aside from this problem, the unfair advantage theory demystifies desert in the relevant respect: it explains that wrongdoers ought to get what they deserve because getting it deprives them of advantages to which they obviously are not entitled. The qualifier "obviously" is important, because it signals that the unfair advantage theory reduces a perhaps superficially mysterious claim (that wrongdoers ought to receive the sanctions they deserve) to an unambiguous one (that wrongdoers ought to receive the sanctions they deserve because receiving them deprives wrongdoers of advantages they obtained unfairly and therefore have no right to retain).

Unlike the unfair advantage theory, the corrective theory does not explain desert's normative significance in terms of something less mysterious. We do not explain why you ought to receive a given punishment for committing a given crime when we assert that, had we threatened you beforehand with the certainty of that punishment, you would have assessed the reasonableness of committing the crime the way an ideal moral agent would assess its reasonableness absent a threat of punishment. To put the matter bluntly, the corrective theory fails the "so what" test. This is a test that the unfair advantage theory unquestionably passes, notwithstanding its other faults. Consider the following bit of dialogue:

- A: D should be sentenced to a year in prison for her robbery.
 B: Why?
 A: Because D's spending a year in prison would deprive her of the advantages she gained unfairly through her wrongdoing.
 B: So what?

B's "so what?" is hard to take seriously. It seems obvious that wrongdoers should not be allowed to retain the fruits of their wrongdoing. The unfair advantage theory passes the "so what" test because it analyzes desert claims in terms of assertions that reasonable people are unlikely to find normatively insignificant—unlikely to ask "so what?" about.

Now consider a different bit of dialogue:

- A: D should be sentenced to a year in prison for her robbery.
 B: Why?
 A: Because if D had known that the robbery would land her in prison for a year, she would have regarded it the same way an ideal agent would: as unthinkable.

- B: Wait, you're saying D *knew* that the robbery would land her in prison for a year?
- A: No, she figured she would probably get away with it. What I'm saying is that if D *had* known that she would get caught and spend a year in prison, she would have regarded the robbery as unthinkable.
- B: Oh. So what?

Here, B's "so what?" is perfectly intelligible. B quite sensibly would like to know why the fact that <threatening D in advance with the certainty of a year's imprisonment for robbery would have caused D to assess the reasonableness of the robbery the way an ideal agent would> entails that D ought to spend a year in prison *now*.

Yaffe's answer to B is in effect his claim that "desert is no more mysterious than similarity" (2018: 103), specifically, similarity between an ideal agent's rational assessment of a wrongful action and a would-be offender's assessment of the same action coupled with a sanction. But the notion of deliberative similarity is mysterious itself, as I argued at the beginning of Sect. 2. And even if it were not, one thing that would remain mysterious is why deliberative similarity is normatively significant—why anything ought to happen to anyone, why anything ought to be done to anyone, *because* any such relation of deliberative similarity exists. Without sounding obtuse, we can ask why a person ought to be made to suffer a given punishment just because she would have reasoned better about what to do had she known in advance that she would suffer that punishment if she acted wrongly. We also can ask whether threatening her in advance really would have led her to reason better, rather than simply to *act* better. By most accounts, people are not reasoning well when they regard wrongdoing as unthinkable because they do not want to get punished for it. People are reasoning well only when they regard wrongdoing as unthinkable *because it is wrong*.

4 Yaffe's affirmative case for the corrective theory

Yaffe's affirmative case for the corrective theory of desert consists of two arguments. Both enhance the theory's plausibility, but neither illuminates desert's normative significance.

Yaffe's first argument defends the corrective theory by showing that it reveals something important about the relationship between desert and culpability, namely, the way differences in mental state yield differences in desert. The argument draws on Yaffe's distinctive conception of culpability: "To be morally culpable for a morally wrongful act is for the act to manifest an imperfection in the way in which one recognizes, weighs, or responds to the reasons generated by the properties of the act in virtue of which it is morally wrongful" (2018: 70).

The argument from culpability goes like this (2018: 117–119):

1. As D's culpability for C-ing increases—from inadvertent risk-taking, to conscious risk-taking, to knowing harm-causing, to purposeful harm-causing—so increases the (inferable) degree to which D fails to recognize, weigh properly, or respond

- appropriately to D's reason(s) not to C. In other words, the more culpable D is for C-ing, the greater D's mode of recognizing, weighing, and responding to C-regarding reasons differs from that of the ideal agent. (Yaffe's conception of culpability)
2. The greater D's mode of recognizing, weighing, and responding to C-regarding reasons differs from that of the ideal agent, the severer a given punishment for C-ing must be in order for the threat of that punishment to induce D to reason about C-ing the way the ideal agent does. (Assumption about human psychology)
 3. If the threat of a given punishment for C-ing would induce D to reason about C-ing the way the ideal agent does, then that punishment is what D deserves for C-ing. (The corrective theory of desert)
 4. Thus, the greater D's mode of recognizing, weighing, and responding to C-regarding reasons differs from that of the ideal agent, the severer a punishment D deserves for C-ing. (From 2 and 3)
 5. Therefore, as D's culpability for C-ing increases, so does D's desert. (From 1 and 4)

When coupled with Yaffe's sensible conception of culpability (Premise 1) and an uncontroversial assumption about human psychology (Premise 2), the corrective theory entails that (arguably, explains why) the extent of an offender's desert tracks the gravity of the offender's culpability.

Although this entailment enhances the corrective theory's plausibility, it does not demystify desert's normative significance: it does not help explain why a person's deserving a given treatment constitutes any kind of reason for the person to receive that treatment. According to Yaffe, highly culpable offenders are those whose deliberations about whether to engage in wrongdoing would have proceeded correctly if and only if they had been threatened in advance with the certainty of severe punishment. But why is that fact a reason to punish these people severely? Why is it a reason to punish them at all? These are questions the argument from culpability does not answer. Nor does the argument from culpability shed light on the *kind* of reason that desert grounds. Does a desert-grounded reason constitute a strict and unyielding demand—a demand of justice, perhaps? Or does it serve instead as a more modest *pro tanto* reason (founded, say, in the supposed intrinsic goodness of wrongdoers getting what they deserve), a reason that we may refrain from acting on when certain other considerations compete? May we thus resist giving the guilty what they deserve if doing so would be extraordinarily costly or would spoil and degrade our moral characters? These are among the questions about desert's normative significance that a theory must answer if it is to demystify desert.³

Yaffe's other affirmative argument for the corrective theory is about the connection between desert and justified punishment (2018: 120–122). This argument seeks to show that the corrective theory can help make sense of why punishments are easier to justify the more deserved they are. The thrust of the argument is that, if the corrective theory is true, showing that a given form of treatment is deserved is a way of undermining a particular kind of complaint about it. Suppose you perform an action that is well-supported by reason. Suppose I proceed to sanction you for the action, and my sanction

³ For more on these issues, see the discussion of desert's "normative force" in Berman 2013.

is so severe that, had you known in advance I would impose it as a certain consequence of your action, the action would not have struck you as reasonable after all. In that case, says Yaffe, you are entitled to complain that I have wronged you by infringing your autonomy. I have retroactively transformed an action of yours that seemed reasonable into one that actually was not. But your complaint will be unavailing if the sanction I impose on you is deserved, Yaffe claims. Per the corrective theory, if the sanction is deserved, your action was one that you never had any right to regard as reasonable. Moreover, the sanction I now impose on you is by definition a sanction the anticipation of which would have led you to reason about the action exactly as you should have. The sanction therefore contradicts an opinion you had no right to hold (that the action was reasonable) and forces on you an opinion you ought to have held anyway (that the action was unreasonable). Thus, by showing that the sanction is deserved, I undermine a complaint you otherwise would have been entitled to make: that the sanction infringed your autonomy by retroactively transforming an action that seemed reasonable into one that actually was not.

Even if sound on its own terms, this suggestive argument does not illuminate the extent of desert's normative significance. Undermining a complaint about a sanction does not show that there is any positive reason to inflict it, let alone demonstrate that inflicting it is justified. In showing that the sanction I impose on you does not infringe your autonomy—in undermining that particular complaint—I do not begin to show that you ought to have received the sanction. Much less do I begin to show that I or anyone else had the authority to impose it.⁴

References

- Berman, Mitchell N. 2013. Rehabilitating retributivism. *Law and Philosophy* 32: 83–108.
- Davis, Michael. 1986. Harm and retribution. *Philosophy & Public Affairs* 15: 236–266.
- von Hirsch, Andrew. 1993. *Censure and sanctions*. Oxford: Clarendon Press.
- Yaffe, Gideon. 2018. *The age of culpability: children and the nature of criminal responsibility*. Oxford: Oxford University Press.

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

⁴ I am grateful to Mitch Berman for comments on a draft of this essay.