RESPONSE TO ADAM KOLBER'S "PUNISHMENT AND MORAL RISK"

*Chelsea Rosenthal**

I. INTRODUCTION

Adam Kolber argues in "Punishment and Moral Risk" that retributivists may be unable to justify criminal punishment, given the reality of moral uncertainty, together with the standards of proof that many of them adopt.[1] And, on his account, moral uncertainty does not pose such stark challenges for consequentialist theories of punishment.

Kolber's argument operates by suggesting that retributivists often take punishment to be unjustified unless high standards of proof are met—we must be quite confident that a defendant actually has committed the crime in question—particularly if punishing the innocent is much worse than failing to punish the guilty. But if retributivists take these commitments seriously, on Kolber's view, they should be applying their high standards of proof to a more general question: are we justified in punishing the defendant? Retributivists may find themselves too uncertain to justify punishing, not only if they have doubts about the facts on the ground, but also, for example, if they have doubts about whether the defendant's actions were genuinely morally wrong in a way that could warrant punishment. Kolber identifies a wide variety of potential sources of doubt for retributivists and concludes that under realistic circumstances they will face too much uncertainty to justify punishing. By contrast, on his account, consequentialists can escape similar difficulties, in part because they do not take there to be a special asymmetry between the badness of punishing the innocent and the badness of failing to punish the guilty. In the face of uncertainty, Kolber suggests, consequentialists will not have to give up punishing because electing not to punish may be comparably risky. And, if retributiv-

* Assistant Professor/Faculty Fellow at the Center for Bioethics at New York University. I am grateful to Adam Kolber for helpful conversation about his paper.

1. Adam Kolber, *Punishment and Moral Risk*, 2018 U. ILL. L. REV. 487.

ists, but not consequentialists, are unable to justify punishing, this raises significant difficulties for retributivism, on Kolber's view. In the later portions of his article, Kolber outlines a way of thinking about cases in which we have some confidence in multiple theories (*e.g.*, consequentialism and retributivism), framing his account in terms of portfolios of beliefs.

In the discussion that follows, I will raise some doubts about Kolber's anti-retributivist argument. I am sympathetic to the idea that we must account for moral uncertainty, and not only factual uncertainty, when determining whether we are justified in punishing. But I am not persuaded that this has the implications for retributivism that Kolber suggests. In Part II, I consider the sources of uncertainty that Kolber takes retributivists to face and argue that several cannot play the role that Kolber's account requires. I reframe his argument in light of these concerns before proceeding to raise more general challenges to it in Part III. Part IV briefly discusses Kolber's notion of portfolios of beliefs.

II. SOURCES OF UNCERTAINTY

On Kolber's account, retributivists face a variety of different forms of uncertainty—uncertainty about the moral features of specific cases (whether a particular defendant's actions were actually wrongful, for example) but also uncertainty about whether retributivism provides the correct account of when punishment is justified.[2] For Kolber, because retributivists have so many reasons for doubting that a particular punishment is justified, it would be inappropriate for them to be highly confident in it, often leaving them unable to justify punishing. But I want to suggest that some of the sources of uncertainty that Kolber identifies cannot do the work that he takes them to do.

Doubts that retributivism is the correct theory of when to punish are fairly different from doubts about whether the standards set up by retributivism are satisfied in a particular case (*e.g.*, because that specific defendant may not morally deserve that punishment). Doubts of the second type fit well into Kolber's argument: We can ask what standard of proof retributivism would apply to these case-specific, moral questions and how likely it is that this standard is met, given our uncertainty. If it is unlikely to be met, then there is a *prima facie* force to Kolber's charge that retributivists will often find punishing unjustified. For example, if retributivists think that punishment is only justified if we are extremely confident that the defendant morally deserves punishment, and we are typically too uncertain to be so confident, this would support Kolber's view.

But, it is not clear how uncertainty about whether retributivism is the correct theory could play a comparable role in his argument—because this involves doubts about whether retributivism's standards are the correct ones in the first place. These doubts do not suggest that punishment will often be unjus-

2. *Id.* at 491–512.

tified under retributivist standards (Kolber's claim); they are just doubts about whether to adopt those standards. Of course, if these doubts are abundant, they might, themselves, provide good reasons to reject retributivism, but this would be independent of Kolber's argument. And, importantly, doubts about the correctness of retributivism do not have to lead the retributivist to doubt whether punishment is justified. Instead, we might have someone who doubts retributivism because they have some sympathies for alternative theories of punishment, under which punishment is *easier* to justify.

Ultimately, too, using general doubts about retributivism to support Kolber's argument has an air of circularity to it. Kolber wants to show that retributivist standards of proof cannot be met, in part because of doubts about the rightness of retributivism. This requires us to apply retributivism's standards of proof to the question of whether retributivism is correct—in order to suggest that those standards are unlikely to be satisfied. But, I am unsure what it means to do this. We would be stipulating that retributivism's standards are correct in order to use them to address whether retributivism is correct. The circularity is made more troubling by the suggestion that the standards would not ultimately be met—that is, that we would not be justified in adopting the retributivism whose standards we were using for the inquiry.[3]

These difficulties for Kolber's argument have two key implications. First, Kolber supports his claim that retributivists cannot justify punishing, partly by providing a list of many different reasons they may doubt whether a punishment is justified. But, if I am correct here, that list should be much shorter. Uncertainty, for example, about whether "suffering (or punishment) is an appropriate response to wrongdoing"[4] raises doubts about whether retributivism is a plausible theory of punishment, rather than doubts about whether to punish a particular defendant, so Kolber cannot rely on this uncertainty without encountering the problems above. There are sources of uncertainty remaining, but removing some forms of uncertainty from consideration means that it will be at least somewhat easier to satisfy retributivists' standards of proof.

Second, this uncertainty about whether retributivism is the correct theory of punishment does not disappear, even though it cannot play the necessary role in Kolber's argument. Instead, it suggests that we should shift our focus. Kolber is correct that the reasonable retributivist will not take the rightness of retributivism to be certain; they will realize that their favored theory might be

3. How worrying this circularity is depends a bit on the precise relationship between retributivism and these standards of proof. If we can characterize the standards of proof as a set of commitments that are separate from retributivism (though often held by its proponents), then the circularity is removed. Kolber does not appear inclined to take this route, given his characterization of the argument as a criticism of retributivism (or of certain retributivist accounts), rather than as an argument that retributivism cannot be combined with particular standards of proof. But, reframing the view in this way may be one option. If the circularity is removed, however, there remains the problem that doubts about the rightness of retributivism do not necessarily decrease our confidence in the appropriateness of punishing (we may doubt retributivism, while having more confidence in alternative theories under which that punishment is justified).

4. Kolber, *supra* note 1, at 489.

mistaken.[5] Once we recognize this, however, it gives us a reason to focus not on what we should do as retributivists (or as consequentialists) but rather on the more general question of how we should behave given uncertainty about the correct theory of punishment. This is the type of inquiry that Kolber takes up in the later portions of his article and that I will return to briefly at the end of this commentary. Shifting from asking what to do as a retributivist to asking what to do as an uncertain person also has the benefit of clarifying the actual implications of doubts about retributivism. Far from giving us a reason not to punish, some doubts about retributivism may give us *more* reason to punish if, for example, these doubts give rise to increased confidence in alternative accounts, under which punishment is more frequently called for.

Before turning to this more general question, however, I want to look more closely at Kolber's challenge to retributivism that receives the main focus of his article. Going forward, I will be taking up a version of his argument that is slightly modified to accommodate the worries above (by focusing more narrowly on determining what retributivism commits us to and deferring doubts about whether retributivism is actually the correct view). The modified claim goes as follows: if we stipulate that retributivism provides the correct theory of when punishment is justified, then punishment is ordinarily unjustified—and if we stipulate that consequentialist theories of punishment are correct, we escape this result. This version of Kolber's view avoids supposing that doubts about retributivism have to lead to doubts about whether to punish, but it encounters other difficulties that I turn to now.

III. HOW RETRIBUTIVISM FARES

It's worth initially highlighting a few key features of this view to get a sense of what it entails. First, this argument does not necessarily require us to reject retributivism. It says that accepting retributivism leads to rejecting punishment. Given that choice, however, we can also opt to reject punishment. Kolber acknowledges this option briefly,[6] but it is worth taking more seriously. In fact, some scholars have argued, for independent reasons, that abandoning punishment is exactly what we should do.[7] If we are retributivists, Kolber's argument could supply us with an additional reason to worry that punishment is often unjustified, rather than a reason to reject retributivism. How we should navigate a tension between retributivism and the justifiability of punishment would depend a great deal on how independently plausible each view seemed, and I will not try to resolve that here. I only want to suggest that we should see the real conclusion of Kolber's argument as a forced choice rather than a par-

5. *Id.* at 528.

6. *Id.* at 491.

7. *See* DAVID BOONIN, THE PROBLEM OF PUNISHMENT (2008); DEIDRE GOLASH, THE CASE AGAINST PUNISHMENT: RETRIBUTION, CRIME PREVENTION, AND THE LAW (2006).

ticular resolution of that choice. But, given that most retributivists do not reject punishment (far from it), this forced choice would itself be an extremely interesting conclusion, if correct.

Second, although Kolber frames his argument as targeting retributivism, in some respects the scope of the argument is both narrower and broader than retributivism. Kolber's argument derives conclusions about the justifiability of punishment from two views it attributes to retributivists—(1) that it is much worse to punish wrongly than to fail to punish, and (2) that we should not punish unless we are very confident that it is justified. Not all retributivists have to accept that punishing wrongly is worse than failing to punish; Kolber discusses Lawrence Solum's worry to this effect.[8] But, perhaps more importantly, it is possible to accept the two views that Kolber targets without being a retributivist. In fact, as Kolber discusses, the values behind these two views are embedded in long-standing legal traditions in the U.S.: in our standard for criminal conviction ("beyond a reasonable doubt" ("BARD")) and in the notion that it would be better to let many guilty men go free than to convict one innocent one.[9] So, in practice, many people do accept these values without being retributivists.

If Kolber's argument is correct, not only will retributivists find it difficult to justify punishment—similar difficulties will be encountered by anyone accepting the two views that Kolber targets. And, because these views are linked to broader BARD values as much as they are linked to retributivism, Kolber's account seems to also pose a challenge to these broader values. But, if there is a tension between these values and punishment, it would be a mistake to presuppose that punishment prevails. BARD values are rooted deeply in our political ideals.

Thankfully, retributivism—and with it, these BARD values—may be able to avoid Kolber's challenge. In turning to the question of whether Kolber's argument prevails, I will follow Kolber and focus primarily on the challenge his view may pose for retributivism in particular, but I take a vindication of retributivism to bring with it a vindication of BARD values that may have faced related difficulties.

So let us ask—is punishment actually unjustifiable, if we adopt a version of retributivism according to which (1) punishing the innocent is much worse than failing to punish the guilty; and (2) we should only punish if we are very confident that it is right to do so? I think this suggestion is on to something important but only partly right. Kolber's key insight is significant—and correct: If we are serious about requiring a high level of confidence before we punish, this should not only mean a high level of confidence that the defendant has engaged in the prohibited activities. Many of the same reasons and values that underlie these demanding standards also suggest we should have a high level of confi-

8. Kolber, *supra* note 1, at 522
9. *Id.* at 502–03.

dence that punishing a defendant is morally justified before we are actually willing to punish them. For Kolber, this leads to doubts about retributivism. But, as I began to suggest above, I am less convinced that this has to be an implication of the insight.

First, Kolber's argument targets retributivism, because, according to the relevant retributivist theories, punishing the innocent is much worse than failing to punish the guilty. Without this asymmetry, forgoing punishment in the face of uncertainty might seem as risky as punishing. Kolber singles out for criticism the forms of retributivism that embrace this asymmetry and offers consequentialist theories of punishment as an alternative that avoids these pitfalls. But, I want to suggest that consequentialism is very susceptible to similar challenges.[10]

In order for consequentialism to avoid this asymmetry, the consequences of punishing the innocent could not be significantly worse than the consequences of failing to punish the guilty. But there are at least some reasons to think that the consequences of punishing the innocent may, in fact, be substantially worse. In part, this is because failing to punish the guilty has impacts that accrue only incrementally. Punishing one innocent person typically has terrible consequences for that person's well-being and the well-being of those close to them. Conversely, failing to punish one guilty person will sometimes produce no particularly bad effects. Many criminals are already unlikely to repeat their crimes, and rarely will leaving one criminal unpunished change general deterrence effects. Even if we look at the effects of policies, rather than individual decisions, it is not at all clear that over-punishing and under-punishing policies have comparably bad effects. Over-punishing has clear and terrible impacts, destroying the lives of individuals and decimating communities. But, what benefits we would miss out on if we under-punished (or punished not at all) is murkier and controversial. Indeed, Deidre Golash has argued that, given the high costs and meager benefits of punishing, we cannot justify punishing on a consequentialist basis (nor, she thinks, on any other).[11] This is, of course, the subject of substantial disagreement that I cannot hope to resolve here. My aim is only to raise doubts about whether consequentialism avoids the asymmetry Kolber sees in retributivism. And, even positions much milder than Golash's will be enough to make it riskier to err towards punishing the innocent than to err towards failing to punish the guilty.

I have also argued elsewhere that, in contexts of moral uncertainty, allowing consequentialism to directly guide our actions raises special risks of wrongdoing.[12] Constraining others' autonomy (*e.g.*, to further the consequences

10. For additional discussion along similar lines, see Lawrence Solum, *Kolber on Punishment and Moral Risk*, LEGAL THEORY BLOG (Jan. 11, 2017, 3:50 PM), http://lsolum.typepad.com/legaltheory/2017/01/kolber-on-punishment-and-moral-risk.html.

11. GOLASH, *supra* note 7, at 22–48.

12. Chelsea Rosenthal, *Why Desperate Times (But Only Desperate Times) Call for Consequentialism*, OXFORD STUD. NORMATIVE ETHICS (forthcoming 2018). Also note that even if we stipulate that consequential-

that seem best) gives a smaller number of agents more of the decision-making power, thereby raising the stakes of possible errors or mistaken moral judgments by those agents. This is risky—especially when there is significant moral uncertainty. Examining how this worry operates in policy contexts (such as the punishment questions at issue here) is outside the scope of this commentary, but it is worth noting as a possible source of further difficulties for the consequentialist.

Finally, whether or not consequentialists encounter similar difficulties, retributivists may be able to weaken Kolber's challenge by appealing to a difference in the standards of proof that are appropriate for particular purposes. According to Kolber, retributivists cannot have enough confidence in the moral justification of a punishment to satisfy something like the BARD standard. But, we should distinguish between the standards we provide in instructions to particular actors within institutions (*e.g.*, jury instructions) and the standards that dictate when the actions of that institution should be seen as morally justified (*e.g.*, when it is justifiable for our legal system to punish). Call the first "instructional standards of proof" and the second "justificatory standards of proof." It is true that many retributivists (and others) believe we should be particularly careful to avoid punishing the innocent, and partly for this reason, they support the use of standards like BARD as instructional standards of proof in relevant institutional contexts. But, this does not mean that they must adopt these standards as justificatory standards rather than taking up other demanding, but less extreme, criteria.

There are good reasons to think that appropriate instructional and justificatory standards of proof may diverge, in part because setting very high instructional standards of proof may help us to satisfy more modest justificatory standards. For example, if a jury's instructions include BARD standards, I usually will not have anything like BARD-confidence that the jury accurately determined the empirical facts of the case; even if the members of the jury have BARD-confidence in good faith, I realize they may be mistaken. But, the fact that a jury was told to follow an extremely high standard (*e.g.*, a BARD standard) may at least help me to have a modest confidence in their conclusions.

When asking whether retributivists must take punishment to be unjustified, we should be asking whether their justificatory—not instructional—standards of proof can be met. The values of many retributivists suggest that we should not punish unless a relatively high justificatory standard of proof is met (especially if, *e.g.*, punishing the innocent is much worse the failing to punish the guilty). But, it is not at all clear that this should mirror the standards they support as instructional standards of proof. And, if the justificatory standards of proof adopted by retributivists are more moderate (if still demanding), there

ism is correct, we may be uncertain about which acts will produce which effects and which effects really manifest the good.

will be fewer cases in which retributivists must regard punishment as unjustified.

So far, then, it seems that if retributivists face difficulties justifying punishment, similar difficulties may be faced more widely, both by consequentialists and by anyone embracing "beyond a reasonable doubt" values. But, retributivists may also have less difficulty justifying punishment than Kolber suggests—some types of uncertainty do not undermine punishment in the way that he proposes, and the appropriate standards of proof may not be as exacting as "beyond a reasonable doubt."

IV. SOME QUESTIONS ABOUT PORTFOLIOS

Near the close of his article, Kolber turns to offering a way for us to think about issues such as punishment, when we are uncertain between theories (*e.g.*, uncertain between consequentialist and retributivist theories of punishment). He suggests that taking into account our varied levels of confidence in competing views can help to make sense of punishment and perhaps other moral topics such as threshold deontology.[13] I am sympathetic to this broad idea. I argue elsewhere that a variety of moral requirements (including threshold deontological requirements) can be better accounted for as ways of mitigating moral risk.[14] But, I have some worries about the particular account that is developed by Kolber. Kolber makes clear that this segment of his proposal is tentative and meant to highlight areas for further exploration. So, my remarks here are primarily intended to raise questions that might be clarified or addressed in future work.

Kolber suggests that we can approach our uncertainty by thinking in terms of "portfolios of beliefs." For example, "[a] tort theorist," in his view, "might be '60% corrective justice-oriented, 40% deterrence-oriented.'"[15] He analogizes these portfolios of belief to stock portfolios:

> Just as one can hold shares of different companies in an investment portfolio, one can hold different beliefs in varying proportions in a portfolio of beliefs. And just as stocks in an investment portfolio interact in ways that can increase or decrease total risk, so too can the constituents of a portfolio of beliefs. Hence, I suggested earlier, retributivism alone might be impotent to punish but capable of doing so if it has a consequentialist backstop. In other words, our backup beliefs should sometimes influence our overall policy preferences.[16]

13. Kolber, *supra* note 1, at 530–31.

14. *See* Rosenthal, *supra* note 12; Chelsea Rosenthal, Ethics for Fallible People (2018) (unpublished Ph.D. dissertation, New York University) (on file with author). Kolber's view was developed independently.

15. Kolber, *supra* note 1, at 529.

16. *Id.* at 529–30.

On this view, combining beliefs well can reduce our risk of moral wrongdoing or help us to navigate difficult moral questions.[17] But, at least on one natural reading, this seems to get the relationship between our beliefs and our choices backward. We may combine financial investments in ways that increase or decrease our total risk, but it does not seem that we can do this with beliefs. First, risk-reduction would be the wrong reason to hold a belief under many epistemological theories. If beliefs should aim at truth, for example, it would be a mistake to select beliefs in order reduce our risk of moral wrongdoing—and, in any case, it is not clear that we could select our own beliefs successfully.

More fundamentally, though, how risky an action is will depend upon the plausibility of moral views that condemn it; we do not adopt beliefs about those views in order to reduce (or increase) the risk. Instead, we manage our risk by adjusting our actions in light of the plausibility of different moral views. For comparison, consider our approach to uncertainty about morally relevant descriptive facts. A consequentialist with 50% confidence that an act will cause pleasure and 50% confidence that it will cause pain should not try to reduce their risk of acting badly by adopting a different combination of beliefs. Kolber may, ultimately, agree with this; this section of his article is more exploratory, and I am unsure how much to make of the metaphor he uses. But these, difficulties do, at least, raise some questions about how the guiding metaphor of a portfolio of beliefs, analogous to a portfolio of stocks, should operate.

V. CONCLUSION

Kolber's project here is an extremely interesting one, and I have enjoyed the opportunity to think more carefully about it. I see one aspect of the account as correct in an important way: The commitments underlying ideas like proof "beyond a reasonable doubt" do seem to support the adoption of high standards of proof for more than just descriptive facts. But, as I have suggested, this does not have the implications that Kolber proposes: It does not give us a reason to reject retributivism.

17. Following Kolber, I use "beliefs" here in the colloquial sense, including both credences of 1, and more limited credences.