

Philosophy and Law



FALL 2014

VOLUME 14 | NUMBER 1

FROM THE EDITORS

Steven Scalet and Christopher Griffin

ARTICLES

Mark Bryant Budolfson

Are There Counterexamples to Standard Views about Institutional Legitimacy, Obligation, and What Institutions We Should Aim For?

Nate W. Olson

Fiduciary Obligations in Medical Research

Katharine Schweitzer

Inquiring into Nature of the Legal Good

Luke Maring

The Truth in Public Reason: Why Retributivism Cannot Make Legal Punishment Legitimate



APA NEWSLETTER ON

Philosophy and Law

STEVEN SCALET AND CHRISTOPHER GRIFFIN, CO-EDITORS

VOLUME 14 | NUMBER 1 | FALL 2014

FROM THE EDITORS

Steven Scalet

UNIVERSITY OF BALTIMORE

Christopher Griffin

NORTHERN ARIZONA UNIVERSITY

In April of 2014, the *APA Newsletter on Philosophy and Law* hosted a conference featuring four emerging scholars working in legal philosophy. This volume contains revised versions of the essays delivered at Northern Arizona University. Mark Budolfson of the Woodrow Wilson School of Public and International Affairs and University Center for Human Values at Princeton University argues against what he takes to be a standard view in political legal theory. This standard view asserts we should aim to bring about the most legitimate feasible institutions and that individual persons have a *prima facie* duty to obey the law. After presenting counterexamples to this standard view, Budolfson argues that there may be no straightforward connection between facts about institutional legitimacy and what reasons, rights, and obligations we actually have or about what institutions we should actually attempt to realize.

Nate Olson, a post-doctoral teaching fellow at Stanford University, investigates the moral and legal relationships between medical researchers and medical research participants, arguing in part that there is reason, contrary to common treatment, to regard the researcher-participant relationship as similar to, but still distinct from, the fiduciary relationship that characterizes the doctor-patient relationship. Olson argues in particular that since participants in medical research are vulnerable in a way similar to patients in doctor-patient relationships, the law ought to regard researchers as having fiduciary obligations to medical participants as do doctors towards their patients even if these duties are not identical.

Philosopher Katherine Schweitzer, University of Nevada, Reno, investigates the cognitive division of labor that tends to separate inquiry into jurisprudence with normative moral, political, and social concerns for justice. In examining questions about the legal good—or the good law ought to do—Schweitzer draws substantially on the scholarship of Robin West, Frederick J. Haas Professor of Law at Georgetown Law, who is concerned that the legal system in which we live has abandoned to a large extent normative theorizing about how law promotes or fails to promote human flourishing. But Schweitzer aims to build on West's thoughts, connecting the project of integrating

legal analysis with normative philosophy West calls for to a pragmatist methodology.

Finally, Luke Maring of Northern Arizona University examines the implications for retributive theories of legal punishment, at least as these aim to provide justifications for punishment at the harsher end of the continuum, of taking seriously the truth in public reason theories of political legitimacy. Maring begins by sharpening up the horns of a dilemma for the retributive theorist offered by Hugo Bedau, and proceeds to refine the public reason requirement that coercive measures by the state be justified to those subjected to the coercion. Maring argues that this requirement varies from law to law, primarily as a function of two factors: the degree of coercion a legal rule threatens against violators, and the presence of adequate but less coercive alternatives. Maring concludes that retributivism cannot satisfy this justificatory requirement for the harsh punishments with which it is traditionally associated.

ARTICLES

Are There Counterexamples to Standard Views about Institutional Legitimacy, Obligation, and What Institutions We Should Aim For?

Mark Bryant Budolfson

WOODROW WILSON SCHOOL OF PUBLIC AND INTERNATIONAL AFFAIRS AND UNIVERSITY CENTER FOR HUMAN VALUES, PRINCETON UNIVERSITY

A standard view in legal and political theory is that, to a first approximation, (1) we should aim to bring about the most legitimate institutions possible to solve the problems that should be solved at the level of politics, and (2) individual people are required to follow the directives of legitimate institutions, at least insofar as those institutions have the authority to issue those directives, and insofar as other considerations are nearly equal.¹ On this standard view, the philosophical analysis of institutional legitimacy can appear to be of the utmost importance because it can seem to answer the most important questions about what we should actually do in the realm of law and politics, and because it may also seem to serve as the crucial bridge between ideal and nonideal theory, at least if institutional legitimacy is analyzed as it often is as a matter of *doing well enough* along dimensions that are familiar from ideal

theory (e.g., ideal justice, universal consent, respecting basic rights, social welfare maximization, and so on).

In what follows I identify a number of possible counterexamples to this standard view, which suggest that there may be no straightforward connection between institutional legitimacy and facts about what reasons, rights, and duties we actually have, and what institutions we should actually try to bring about, contrary to what the standard view assumes.

SHOULD WE AIM FOR LEGITIMATE INSTITUTIONS?

Consider the following view about what institutions we should try to bring about, which might be called the *Aim for the Most Legitimate Feasible Institution View*:

First, identify the set of institutions that are *feasible*; then, rank those institutions by *legitimacy*. The institution we should actually try to bring about is the most legitimate feasible institution.

This view is mistaken because it ignores the fact that it might be unacceptably risky to aim for the best feasible institution. As an initial toy example to illustrate, suppose that an asteroid will destroy the planet unless the world cooperates to stop it, and suppose that there are only two feasible solutions, A and B. Solution A is slightly more legitimate than B, but the probability that we will succeed if we aim for the more legitimate solution A is only 90 percent, whereas it is 99 percent certain we will succeed if we aim for the slightly less legitimate solution B. In such a case, it could easily be true that we should aim for the less legitimate solution and its associated institutions, on the grounds that the expected outcome of pursuing the less legitimate solution is sufficiently better in all respects than the expected outcome of pursuing the more legitimate solution, even though the more legitimate solution may well be feasible in light of its sufficiently high likelihood of success.

This objection can also be illustrated by real-world cases in which there is no emergency in play. For example, suppose that overfishing threatens many species of fish that travel through international waters, and to curb it we need an effective international treaty that creates the right kind of incentives to ensure more sustainable fishing. Suppose that in order for such a treaty to succeed, we need a critical mass of nations to join as initial compliant nations and impose penalties on nations that do not comply, leading over time to more and more nations complying with the treaty.² In such a case, a simple treaty that ignores many of the interests of developing nations might have a much higher probability of success than a slightly more legitimate and also feasible treaty that includes provisions to subsidize developing nations for the costs that the treaty will impose on them that ideally they should not be expected to bear. However, if the probability of success of the simpler and slightly less legitimate treaty is sufficiently higher than the probability of success of the slightly more legitimate redistributive treaty (because the developed nations might have trouble selling the idea of subsidizing

the economic activity of developing nations to their legislatures, and thus might have trouble securing as quick or as smooth of ratification of the slightly more legitimate treaty), and if the costs to the developing nations of the slightly less legitimate treaty are not onerous, and if the costs of a failure to adopt any treaty are high (because that would mean additional decades of overfishing), then we should introduce and promote the slightly less legitimate treaty, even though there is another treaty available that is both feasible and more legitimate. Once again, this shows that if we must choose which institution to aim for, and one is more legitimate than the other feasible alternatives, it does not follow that we should try to bring about the most legitimate feasible institution.

The upshot is that we will often go wrong if we rely on the common Aim for the Most Legitimate Feasible Institution View, because that view ignores the probabilistic complications of real-world decision making. At the very least, what we need instead is a familiar sort of decision theory that takes into account facts about the expected probability of success, failure, unintended consequences, and so on that could result from attempts to bring about the various possible institutions that we have to choose from—and sometimes the institution that such a decision theory will say that we should try to bring about will be less legitimate than other feasible institutions, contrary to what theorists who discuss institutional legitimacy tend to assume. As this makes clear, the point here is not about the legitimacy of the *decision procedure* we should use in deciding what course of action to choose—for all I've said, it remains true that we should always use the most legitimate *decision procedure*. The current point is not about the *decision procedure*, but rather about the *goal*—the current point is that our goal should not always be merely to bring about the most legitimate feasible institution.³

In light of this objection to the Aim for the Most Legitimate Feasible Institution View, theorists might drop that view and instead adopt what might be called the *Maximize Expected Legitimacy View*:

First, evaluate all possible institutions with respect to legitimacy; then, figure out the probability of success of aiming for each of those institutions, etc.; then decide which institution to aim for in a way that *maximizes expected legitimacy*.

Although somewhat vague, this idea is clear enough for our purposes because it gestures at the sort of view that theorists may initially find most attractive. Basically, the idea is to adopt a familiar sort of decision theory, where the point of this particular decision theory is to identify the course of action that can be expected to best promote the aim of institutional legitimacy given the real-world facts.

Unfortunately, this view is also mistaken, because the values that are relevant to institutional legitimacy are not the only values that we should care about, even from a “purely political” point of view; as a result, the course of action that best promotes the aim of institutional legitimacy might fail to be the course of action we should actually take, because it might fail to best promote the broader

set of values we care most about even from a “purely political” point of view.⁴ As an example, imagine again a scenario in which overfishing threatens some fish stocks in international waters, but where this time if nothing is done the fishery collapse will devastate the poor nations in the immediate vicinity and will also cause some lesser harm to other nations around the world. Fortunately, although it will be quite costly to solve the problem, it can certainly be done. Furthermore, the poor nations in the immediate vicinity of the impact are willing to pay the entire cost of solving the problem because of the even more important things they stand to lose. So, one possibility is for the rich nations to stand by as free riders and simply let that small group of mostly poor nations pay the entire cost of the solution, which will then benefit the rich nations at no cost to them. However, although such an outcome and its associated institutions would be legitimate because everyone and all nations would *freely consent* to them in the right sort of way, it would by no means be the best feasible course of action all things considered, because a consequence would be that most of the world would free-ride on an important and expensive global public good in a way that is very unfair, especially given that very poor nations would pay most of its cost. Instead, we can imagine that there is a way to secure a better much outcome, although it is less legitimate because it requires a powerful nation such as the United States to threaten and cajole other nations into contributing to the public good in a way that is, as far as such things go, fairly benign, but which also makes the consent of those other nations not freely given—and we can imagine that such a power play also favors American interests in a way that is itself slightly objectionable. Nonetheless, although we can suppose that the result of such a power play would be less legitimate because of the lack of full and freely given consent and the slight favoritism shown to the most powerful nation, it might easily be preferable because it would be much more fair on balance than the more legitimate institutional outcome in which the small group of nations is allowed to bear the entire cost of a global public good themselves.

As this case shows, even when the probabilities of success are held constant, sometimes we should aim for institutions that are less legitimate. As a result, there appear to be counterexamples to the idea that we should always act to maximize institutional legitimacy, and more generally to the idea that there is a straightforward connection between institutional legitimacy and what institutions we should try to bring about.⁵

IS THERE A PRIMA FACIE DUTY TO OBEY THE LAW?

The arguments so far suggest that the notion of institutional legitimacy may not be the key to explaining what institutions we should actually try to bring about. In what follows, I argue that there also may not be a straightforward connection between institutional legitimacy and what political obligations we have—that is, what rights and duties individuals have with respect to the directives of nonideal institutions. This is contrary to what is surely the standard view, which might be called the *Prima Facie Duty to Obey the Law View*:

If an institution is *legitimate* (and/or has *the authority to issue directives* and/or has *the right to rule* and/or has been *freely consented to by everyone*), then individuals have content-independent moral reasons to obey its directives (at least when those directives are themselves legitimate).⁶

Here is the general form of a counterexample to this view: Imagine that a particular institutional arrangement is the uniquely reasonable solution to a real-world problem, and that in order to function efficiently the institution must assert authority over and issue directives in its domain to everyone. At the same time, such an institution must offer mistaken directives in some localized areas in a way that is easily foreseen (including by those locals), because it would clearly introduce a very counterproductive amount of complexity into the institution if it tried to take these particular local subtleties into account. Such a case is possible because, for example, a global problem might require consent of everyone to an optimal-in-practice set of directives, which are known to give mistaken verdicts in a range of cases that local individuals can readily identify, where those directives cannot be improved upon because of feasibility constraints, such as information gathering and processing constraints at the institutional level, too great of expense associated with a more subtle set of directives, or that giving discretionary powers to locals to override the simple directives when appropriate would be predictably misused in too many cases, thereby undermining the global solution. In such a case, everyone might be required to consent to such an institution and its directives, which might well be sufficient to make the institution a legitimate authority and all of its directives legitimate, but yet locals who are issued the foreseeably mistaken directives might well have no “content independent reason” to follow those directives, and so would have no *prima facie* duty to obey those directives, contrary to the *Prima Facie Duty to Obey the Law View*.

As an initial toy example to illustrate, imagine that we build a sophisticated system to manage our nation’s transportation system that will benefit everyone greatly. For example, it will provide near-perfectly efficient management of traffic signals, which will generate additional economic growth and leisure time, lower accident rates, and so on. However, as the system is in its final test stages, a computer glitch emerges that implies that the system will offer many mistaken directives at a few specific locations each leap year on February 29. Furthermore, because the software and the hidden bug are so complex, it will likely take many years to develop a superior replacement for this near-perfect system, and there is no practical way of preventing the system from issuing the localized mistaken verdicts every February 29, because this would undermine the optimal directives that it will still issue everywhere else on that date. All of this information is conveyed to the public, and a referendum is held to decide whether to implement this system despite the glitch, and everyone votes enthusiastically in favor of implementing the system and making it the centerpiece of the nation’s transportation-system governance institution, including those who live in the few locations where it will issue many

mistaken directives on each February 29. As a result, the ensuing institution and its directives are clearly legitimate. Nonetheless, in the few localities where its directives are known to be completely unreliable on February 29, individuals there have no “content-independent reason” to obey its directives issued on that date—instead, they should form their own judgment, without consulting the directives of the legitimate institution at all. In other words, it is not as if those locals need to recognize any *prima facie* reason to follow the directives of the traffic lights in their locality on February 29 in a way that can then only be overridden by conclusive evidence that the directives of those traffic lights are mistaken. On the contrary, it would be a mistake for them to reason in such a way (e.g., it would simply be a mistake to think that those traffic signals provided anything like “exclusionary reasons” on that date in those locations), and instead locals should simply ignore the traffic signals entirely on that date, and make decisions directly on the basis of the non-institutional reasons as they judge them. (In a longer version of this paper, I suggest, based on further arguments, that there may be no such thing as the sort of “content independent,” “exclusionary” reasons commonly cited in connection with these issues.⁷)

This sort of case suggests that, contrary to the *Prima Facie* Duty to Obey the Law View, the directives of an institution do not necessarily create special reasons to obey even if that institution is fully legitimate, has the right to rule, has the authority to issue directives, has been freely consented to by everyone, and issues directives that are themselves legitimate.

For a more realistic example, imagine again a world in which a new institution is necessary to prevent overfishing of the oceans, where some small fishing villages realize that such an institution, although necessary, will inevitably issue mistaken and counterproductive directives to their small corner of the ocean for the sort of basic practical reasons indicated above. In such a case, if the institution is necessary to avoid a global calamity, everyone—including those locals—might have sufficient reason to vote in favor of its creation—and could even have a moral obligation to do so—but yet its directives might not give them any content-independent moral reasons, and so they might not be under any obligation to comply with its directives even if they and the entire world freely consented to the institution knowing it would issue exactly those directives.⁸ Such a case is possible, again, because a global problem might require consent of everyone to an optimal-in-practice set of directives, which are known to give mistaken verdicts in a range of cases that local individuals can readily identify, where these directives cannot be improved upon because of feasibility constraints, such as information constraints that the institution faces, too great of expense associated with a more subtle set of directives, or that giving discretionary powers to override the simple directives would be predictably misused in too many cases, thereby undermining the solution.

In evaluating the significance of these examples, it is important to see the crucial disanalogies between these cases and the familiar stoplight in the desert case where a person permissibly runs through a traffic light in the

middle of the desert that directs her to stop because she can see that there are no other cars around for miles. In this stoplight in the desert case, it is plausible to argue that the driver has some special reason to stop at the light simply because the light directs her to stop, but that that reason is *outweighed* by the other considerations that are present in the case. As a result, the stoplight in the desert case does not provide a counterexample to the *Prima Facie* Duty to Obey the Law View, because it is easy to maintain that the driver has some important *prima facie* duty to obey this traffic light regardless of its signal, but that that *prima facie* duty is outweighed in this particular case.

In contrast, in the examples provided here there is no clear way of maintaining that the locals are under any special obligation to comply that needs to be outweighed in the situations described. Instead, in the situations described here, the directives of the institution provide them, in their idiosyncratic location, with *no positive reasons to obey*, despite their freely given consent to such an institution and its directives, including knowledge that it would issue *these very directives*. In other words, while it is plausible to claim that in the stoplight in the desert case reasons to obey are present but outweighed, in contrast in the cases described here there appear to be no reasons to obey at all and thus no reasons that need to be outweighed. As a result, the cases here appear to provide counterexamples to the *Prima Facie* Duty to Obey the Law View, and thus go well beyond familiar examples such as the stoplight in the desert case in which obligations to obey the law are outweighed by other important considerations.⁹

CONCLUSION

In sum, I’ve identified a number of objections to a standard view in legal and political theory according to which, roughly, (1) we should aim to bring about the most legitimate feasible institutions, and (2) individual people have a *prima facie* duty to obey the law. Contrary to this standard view, the examples here suggest that there may be no straightforward connection between, on the one hand, facts about institutional legitimacy and facts about what reasons, rights, and duties we actually have and what institutions we should actually try to bring about, contrary to what the standard view assumes. These results threaten the idea that the analysis of institutional legitimacy is of particular importance because they threaten the idea that institutional legitimacy is the key to answering important questions in and about non-ideal theory.

* For a more in-depth discussion of these and related issues, see my paper, “Questioning the Relevance of Legitimacy and Authority to Nonideal Theory.” Thanks to John Devlin, Chris Griffin, Luke Maring, Kevin McGravey, Nate Olson, Steve Scalet, and Katharine Schweitzer for helpful comments.

NOTES

1. See, for example, Allen Buchanan and Robert Keohane, “The Legitimacy of Global Governance Institutions,” *Ethics and International Affairs* 20, no. 4 (2006): 405–37; Thomas Christiano, “The Legitimacy of International Institutions,” in *The Routledge Companion to Philosophy of Law*, ed. Andrei Marmor (New York: Routledge, 2012); A. John Simmons, *Justification and Legitimacy* (New York: Cambridge University Press, 2001); David Estlund, *Democratic Authority* (Princeton, NJ: Princeton University Press, 2009); Christopher Heath Wellman, “Samaritanism and the Duty to Obey the Law,” in *Is There a Duty to Obey the Law (For and Against)*, eds. Christopher Heath Wellman and A. John Simmons

(New York: Cambridge University Press, 2005); John Rawls, *A Theory of Justice*, rev. ed. (Cambridge, MA: Belknap Press, 1999); John Locke, *Second Treatise of Government* (1690); and many others who endorse one or both of (1) and (2).

2. For discussion of structurally similar treaties to combat global environmental problems, see Scott Barrett, *Environment and Statecraft: The Strategy of Environmental Treaty-Making* (Oxford University Press, 2003), especially chapter nine. For discussion of such a treaty in connection to climate change, see Joseph Stiglitz, "A New Agenda for Global Warming," *The Economists' Voice* 3, no. 7 (2006); and my paper "Global Justice, Political Realism, and the Ethics of Collective Action," Available online at <http://www.budolfson.com/papers>.
3. As a result, my discussion here does not apply to "more theoretical" discussions of legitimacy that have no clear upshot for what we should actually do in the real world, such as John Rawls's discussions and many other ideal theory discussions. My goal in this paper is to engage with views in legal and political theory that have implications for what we should actually do about the world's problems.
4. Whatever "purely political" is taken to mean—here I have in mind "public reason" views, and so on.
5. An important consequence of all this is that even when the probabilities of success are held constant, sometimes we should prefer an institution that is *not legitimate* to one that is legitimate. To see how this result follows, consider the case just described in which we should prefer a less legitimate institution (call it "B") to a more legitimate institution ("A"), even though the probability of success of both institutions is the same. Now interpret the case so that A is ever-so-slightly above the threshold for legitimacy simpliciter while B is ever-so-slightly below; in that case, we should still shoot for the institution that is more fair but not legitimate (B) rather than the institution that is legitimate (A) even though the probabilities of success are the same.
6. See, for example, John Rawls, *A Theory of Justice*, rev. ed., (Cambridge, MA: Belknap Press, 1999), 312 (Section 53); Allen Buchanan and Robert Keohane, "The Legitimacy of Global Governance Institutions," *Ethics and International Affairs* 20, no. 4 (2006): 411; Thomas Christiano, *The Constitution of Equality* (New York: Oxford University Press, 2008), 240-41; A. John Simmons, *Justification and Legitimacy* (Cambridge, MA: Cambridge University Press, 2001), 106 and 130; David Estlund, *Democratic Authority: A Philosophical Framework* (Princeton, NJ: Princeton University Press, 2009), 154-155; Christopher Heath Wellman, "Samaritanism and the Duty to Obey the Law," *Is There a Duty to Obey the Law (For and Against)*, eds. Christopher Heath Wellman and A. John Simmons (New York: Cambridge University Press, 2005), pp. 52-53; John Locke, *Second Treatise of Government* (1690), and many others. For our purposes, we can understand the *Prima Facie* Duty to Obey the Law View as agnostic on the issue of whether the relevant duty is owed to the state or to the citizens of the state.
7. See my paper, "Questioning the Relevance of Legitimacy and Authority to Nonideal Theory" (unpublished manuscript). Thanks to John Devlin for particularly helpful discussions here (I don't know whether or not he would agree with my commentary about this example).
8. In fact, despite having freely consented to the institution, those locals would not necessarily have an obligation to show any interesting kind of respect to the institution at all—for example, they might foresee that in all of their interactions with the institution in their corner of the world the thing to do will be to actively interfere with its operations, by moving monitoring equipment and marker buoys, etc., because in their corner of the world the institution's directives will be reliably misguided in a way in which the best redress is clearly to interfere in such a way. (Here I disagree with the view defended in Buchanan and Keohane, "The Legitimacy of Global Governance Institutions," see especially page 411.)
9. For other familiar examples that are also consistent with the *Prima Facie* Duty to Obey the Law View, see Kent Greenawalt, *Conflicts of Law and Morality* (New York: Oxford University Press, 1987).

Fiduciary Obligations in Medical Research

Nate W. Olson

STANFORD UNIVERSITY

1. INTRODUCTION

Traditionally, the roles and obligations of medical clinicians and researchers have been thought to be quite distinct. Clinicians' primary objective is to look after the health of their patients, and, as a result, both professional organizations, such as the American Medical Association, and courts have held that the physician-patient relationship is a fiduciary one.¹ In a fiduciary relationship, one individual, the entrustor, gives another individual, the fiduciary, the authority to look after some aspect of his or her well-being. The fiduciary must then use that authority to act in the best interest of the entrustor. In a physician-patient relationship, a patient entrusts his or her health to his or her physician, and the physician bears a responsibility to always act in the patient's best interest. A physician therefore clearly bears the marks of a fiduciary.

Medical researchers, on the other hand, are commonly thought not to be fiduciaries for the participants in their trials. After all, the point of medical research is not to benefit a trial's participants, but to produce generalizable knowledge. Consider, for instance, that many trials of new medical treatments are placebo trials, where participants in one arm of the trial receive the experimental treatment being tested and those in the other receive a placebo. While study participants often harbor the "therapeutic misconception," the view that a trial is intended to benefit them, the use of placebos makes trials' true purpose clear: to produce scientific data about the effectiveness of a treatment.

It is understandable then that the common view in research ethics has been that while researchers have obligations to inform their participants about the details of trials and not expose them to excessive risks, they have very limited obligations to provide benefits to their study participants. Provision of some benefits may be required to achieve an acceptable risk/benefit ratio for participants and so protect them from being exposed to excessive risks.² But a researcher's obligation to benefit her participants stops at such protections. A researcher's primary responsibility is to follow her study's protocol, not to provide such benefits. Unlike physicians, researchers do not seem to be fiduciaries.

However, some ethicists have begun to argue for a more capacious understanding of researchers' obligations. Two developments in medical research have helped spur this turn away from the traditional view. First, as diagnostic tools have continued to improve, the prevalence of incidental findings in research is dramatically increasing. For instance, in studies involving CT colonographies, which produce an image of one's colon and the surrounding area, researchers find conditions outside the colon of potential medical significance in approximately 10 percent of study participants.³ Ethicists are increasingly arguing

that researchers have a responsibility to at least report these results back to research participants and, in some cases, such as when research is conducted in a low-income country and the cost of treatment is low, even provide treatment for these conditions themselves.⁴ Second, researchers are increasingly making use of biobanks (i.e., repositories of biological samples). After biological samples are taken from participants for use in one study, they are often sent to a biobank so that researchers can use them for future studies. When conducting subsequent studies, researchers are increasingly discovering medically significant conditions. Here, too, ethicists have argued that researchers have an obligation to report results back to the person who donated the sample.⁵

Even if we adopt this more capacious understanding of medical researchers' responsibilities, as I think we should, their responsibilities remain distinct in character from physicians' responsibilities in a few ways I detail in this paper. They therefore fall somewhat short of full-fledged fiduciary obligations. Yet the researcher-participant relationship also does not fit into other categories of relationships commonly recognized in the law, such as contractual ones. In this paper, I argue that the law ought to recognize that researcher-participant relationships occupy a space close to but distinct from fiduciary ones. By examining how researcher-participant relationships are both distinct from and similar to fiduciary relationships, my argument helps us better understand researchers' ethical and legal obligations to their participants.

To see how we ought to understand researchers' obligations, I begin in the next section by identifying four main features of fiduciary relationships. In section three, I explain the extent to which the researcher-participant relationship exhibits each of these features and argue that the same rationale for fiduciary obligations carries over to the researcher-participant relationship. In particular, fiduciary obligations stem from the vulnerability that entrustors incur by entering fiduciary relationships, and, I argue, this same kind of vulnerability exists in the researcher-participant relationship. In section four, I show that this rationale extends to research involving biobanks before concluding in section five that although there should be some legal duty, the law ought to resist the temptation to force the researcher-participant relationship into a pre-existing legal category.

2. FIDUCIARIES

The concept of a fiduciary relationship stems from legal discussions of trusts, and the entrustor-trustee relationship remains the paradigmatic example of a fiduciary relationship. In a trust, an entrustor places property in the care of a trustee for the benefit of a third party, the beneficiary. Over time, the law has broadened its understanding of fiduciary relationships beyond cases involving property. Individuals may also be regarded as fiduciaries if they represent another's interests or act in an advisory role where the entrustor must rely on their advice or expertise.⁶ In these cases, the entrustor and the beneficiary can be the same individual, as in the physician-patient relationship. (Since the entrustor and the beneficiary are the same individual in the cases that are our focus here, I will simply use the term "entrustor" to refer to this individual.) Although pinning

down a precise definition of fiduciary relationships has proven difficult,⁷ four characteristics unite these different types: transfer of power, discretion, mutual voluntariness, and duties to use the power for the entrustor's best interest. Fiduciary relationships share the mutual voluntariness feature with contractual relationships, but the other three features distinguish them. The first two features combine to create a relationship that is distinctly unequal, and the last one, the duties to use the power for the entrustor's best interest, serves to protect the entrustor from the fiduciary misusing her superior position.⁸

First, the basic idea of a fiduciary relationship is that one individual gives another the power to look after an item or an aspect of the individual's welfare. This transaction removes that item or that aspect of the individual's welfare from the entrustor's care and passes it to the fiduciary. This creates a relationship unequal in two respects. First, and most obviously, the entrustor, as the name indicates, now is in a position where she trusts the fiduciary to look after the item or aspect of her welfare. In cases involving advisors, such as the physician-patient one, the fiduciary is given this power due to the specialized knowledge she possesses. Having a lower level of expert knowledge puts one in an unequal position vis-à-vis a physician, but becoming a physician's patient introduces a different sort of inequality. Namely, a patient trusts his physician to provide him with medical care and advice. If the physician does not tell him he has symptoms of heart disease, he may think that he does not have such symptoms. However, if he sits next to a physician on a bus and the physician does not tell him that he has symptoms of heart disease, he may not draw the same conclusion. We give our physicians the responsibility to look after our health, and so we trust them to do so. It is not just that someone is a physician that is the basis of the trust; it is that she is *our* physician.

While most patients do trust their physicians to look after their health, importantly, it is not the actual trust that forms the basis of the relevant inequality. Rather, it is that patients are *entitled* to trust their physicians, whether or not they in fact do.⁹ *Entrustment* lies at the heart of the fiduciary relationship, not the psychological state of trust. By transferring the power to look after one's welfare to a fiduciary, one entrusts this power to him or her. Actual trust often follows in the wake of this transfer, but it need not. So, one way that fiduciary relationships are unequal is that they involve a one-way transfer of power to look after an item or aspect of one's welfare.¹⁰

The second sort of inequality is due to the fact that fiduciaries have the responsibility to look after someone else's welfare or property. I might gift you my prized baseball card collection and trust you to take good care of it. But after gifting it to you, the collection is no longer mine. On the other hand, if I entrust it to you for safekeeping while I travel around the world for a few months, I retain ownership.¹¹ So, it's not just that one transfers a power to another that makes a fiduciary relationship unequal, it is that one gives another the power to look after one's own welfare or property. The distinction between gift and fiduciary transactions will be important for my discussion of biobanks in section four.

Second, fiduciaries possess discretion to decide how best to act in the entrustor's best interest. Possessing such discretion is part of what the power to act in another's best interest involves. When looking after an individual or an item's welfare, unexpected circumstances are likely to arise, and fiduciaries must be allowed to use their best judgment to determine how to act in light of such circumstances.¹² The fact that fiduciaries possess such discretion adds another layer of inequality to the relationship. Although, as I will explain below, a fiduciary is duty-bound to act in his or her entrustor's interest, the entrustor cannot precisely specify all the ways that the fiduciary ought to act. The relationship is therefore unequal in this respect even if the fiduciary fulfills his or her duty to act in the entrustor's best interest.

Third, fiduciary relationships are formed through a mutually voluntary transaction. The entrustor offers the fiduciary authority over an item or aspect of her welfare, and the fiduciary accepts the offer. One does not possess fiduciary obligations simply because another individual relies on his judgment. For instance, perhaps my daily morning routine involves looking out of the window to see whether my punctual and responsible neighbor is toting an umbrella as she walks out the door. If she brings an umbrella, I bring one, too; if she doesn't, I don't. Though I rely on her to stay dry, it's clear that she bears no obligation to act in my best interest. If she started hiding her umbrella in her backpack when she walked out the door, I would have no grounds for complaint. That both parties agree to the transaction, then, is crucial for the establishment of a fiduciary relationship.¹³

Finally, a fiduciary must use the power with which she has been entrusted for the benefit of the entrustor. Often, fiduciaries are characterized as possessing twin duties of loyalty and care.¹⁴ They have an obligation to care for the entrustor—that is, use their entrusted authority to act in the entrustor's interests—and they must remain loyal to the entrustor's interests even when presented with opportunities to advance their own or third parties' interests at the expense of the entrustor. For physicians, this means that they must make medical decisions on the basis of what's best for their patients, even when this conflicts with, for example, a patient's employer's wishes or maximizing profits.

The duties of care and loyalty mark a clear distinction between fiduciary relationships and contractual ones. In a well-known characterization of fiduciaries, Judge Benjamin Cardozo writes,

Many forms of conduct permissible in a workaday world for those acting at arm's length are forbidden to those bound by fiduciary ties. A trustee is held to something stricter than the morals of the market place. Not honesty alone, but the punctilio of an honor the most sensitive, is then the standard of behavior.¹⁵

Unlike contractual relationships where individuals are permitted, and even expected, to act in self-interested ways so long as their actions do not violate the terms of the contract, fiduciaries must always use the power afforded them for the benefit of the entrustor. Moreover, fiduciaries'

duties of care and loyalty cannot be negotiated away.¹⁶ Fiduciaries bear them even if they are not recognized by the terms of a contract.

Fiduciaries are held to a higher standard than contractors because of the trust that is placed in them and entrustors' subsequent vulnerability to the decisions they make. The paradigm for contractual relationships is that of two equally placed, self-interested individuals who make an agreement for mutual benefit. Fiduciary relationships, on the other hand, are fundamentally unequal. Even if individuals are relatively equally placed before entering into a fiduciary relationship, we have seen that the structure of fiduciary relationships is such that once the relationship is established, an entrustor stands in an unequal position to his or her fiduciary. This inequality puts entrustors in a vulnerable position, in the sense of standing in threat of harm or mistreatment.¹⁷ They are particularly vulnerable to exploitation, as when the fiduciary uses the transferred power to advance his or her own interests,¹⁸ but they are also vulnerable to other forms of misconduct, such as simple negligence (of the duty to act in the entrustor's best interest).¹⁹ The duties of care and loyalty are in place to protect against such misconduct.

In sum, then, we can say that a fiduciary relationship begins with a mutually voluntary agreement for an entrustor to transfer the power to act in his or her best interest to a fiduciary. Once the relationship is instituted, duties of care and loyalty bind the fiduciary to act in the entrustor's best interest.

3. THE RESEARCHER-PARTICIPANT RELATIONSHIP

We are now in position to assess whether researcher-participant relationships are fiduciary ones. As I have already stated, the main objection to classing them as fiduciary is based on the fact that researchers' primary responsibility is to produce generalizable knowledge, not benefit their participants.²⁰ Earlier, I used the example of placebo arms of trial to illustrate this point, but in many placebo trials, there are no existing proven treatments for the medical issue under study. So while a placebo does not provide benefit (other than the placebo effect), those who receive placebos in such cases are not being blocked from receiving the standard of care (i.e., the normal course of treatment). However, research participants are sometimes precluded from receiving the standard of care for other reasons. For instance, some trials involve a "washout" period in which participants stop taking previous medications. While such washouts are not allowed if the risks to participants are too high, this example shows that researchers may sometimes provide sub-optimal care to their participants in the service of their primary goal of producing generalizable knowledge. This sort of example, even more than the placebo one, casts doubt on the claim that researchers are fiduciaries.²¹

In fact, I believe this kind of example shows that researchers are not full-fledged fiduciaries. However, as I will argue in this section, they do possess several similar features to fiduciaries and so ought to be treated similarly by the law. In particular, I argue that through granting researchers

access to private health information, participants become vulnerable to researchers' actions in much the same way that entrustors are vulnerable to fiduciaries.

The first step to addressing the above objection is to note that possessing an obligation to produce generalizable knowledge does not preclude researchers from also possessing a quasi-fiduciary obligation to their participants. Even if their obligation to produce generalizable knowledge trumped in every case of conflict, researchers still could be obligated to benefit their participants whenever such a conflict does not exist. But it's clear that the obligation to produce generalizable knowledge does not trump in every case of conflict. Researchers are prohibited from exposing participants to excessive risk even if so exposing them could further advance the goals of their research. So, researchers must keep the interests of their participants in mind to some extent. That said, an obligation to avoid exposing participants to excessive risk is an obligation to avoid harming them, not an obligation to benefit them. It's therefore not the same as a fiduciary's obligation to care for her entrustor.

To establish that researchers do have such an obligation, we need to look to the underlying rationale for the obligation of care in fiduciary relationships: the vulnerability of entrustors. By looking at the issue of incidental findings, we can see that research participants are subject to the same sort of vulnerability. For example, an echocardiogram, a type of heart scan, conducted as part of a study on the effects of severe malaria on pulmonary function may show that a participant has a heart infection.²² Assume that the participant does not have other means of knowing that she has this heart infection, which, unless treated, can lead to life-threatening complications.²³ She is therefore in a quite vulnerable position. If the researcher reports this information back to her, she can begin to seek treatment. If the researcher does not report it, she may begin to develop severe health problems.

But this is not yet enough to establish that participants face the same type of vulnerability as entrustors, that is, the vulnerability that comes from giving fiduciaries discretionary power to look after an aspect of their welfare. As I explained in the previous section, entrustors are entitled to trust that fiduciaries will carry out this power and are vulnerable to them actually doing so. Patients are entitled to trust their physicians to act in their best interest because this is the agreed-upon purpose for entering into a physician-patient relationship, a purpose codified by physicians' professional code and relayed to patients in official documents. Yet, as we have seen, this is not the purpose of the researcher-participant relationship. Therefore, *prima facie*, it may seem that participants are not entitled to trust researchers to act in their interest and so lack the relevant sort of vulnerability.

The key to seeing why participants are in fact entitled to trust researchers to act in their best interest rests with the agreement that happens at the outset of the researcher-participant relationship: the agreement to participate in research. This agreement follows the requirements of the informed consent process, including the researcher

stating the research procedures and the participant agreeing to these procedures. The participant thereby permits the researcher to perform procedures that would be impermissible absent such an agreement, for example, drawing blood and taking an echocardiogram. Henry Richardson and Leah Belsky have argued that through granting such permissions, participants "partially entrust" their health to medical researchers.²⁴ On their account, researchers bear obligations to look after the aspects of participants' health that are made known to them by following the authorized procedures. For instance, since the heart infection is discovered via one of the study's procedures, the researcher has an obligation to look after this aspect of the participant's health, where this might mean providing treatment for the infection herself or referring the participant to someone else.²⁵

Richardson and Belsky point out that entrustment relationships come in several forms; paradigmatic fiduciary relationships represent one extreme where the entrustment is much fuller than in their "partial-entrustment model." Rather, they appeal to a more limited form of fiduciary relationship called a "bailment," in which someone "accepts custody of some particular good and is entrusted to look out for it only in limited ways."²⁶ Importantly, bailment relationships can be undertaken for other purposes than transferring power to look after an item's good. Their example is of an auto shop that takes custody of one's car in order to fix a dent in it. Although this objective is very limited, the auto shop incurs a broader responsibility to look after the car. For example, it can have a responsibility to move the car when a fire breaks out next door.²⁷ Richardson and Belsky draw an analogy to researcher-participant relationships, which are undertaken to produce generalizable knowledge, but nevertheless, they argue, contain bailment responsibilities.²⁸

While the auto shop analogy does show that a fiduciary relationship can exist even when a relationship is entered into for other reasons, the analogy is strained in one respect. Participants do not hand over pieces of property when agreeing to enter trials. Rather, they grant researchers permission to perform medical procedures on their body (which the law treats as one's person, not one's property) and access private health information. Because of this disanalogy, Richardson has removed reference to bailment in his more recent work on this subject and instead now focuses on the fact that such permissions involve waivers of privacy rights.²⁹ He uses the waiver of privacy rights as the basis for a moral argument that researchers possess obligations to look after the welfare of their participants.³⁰

Given the disanalogy and his focus on the moral basis of researchers' obligations, it makes sense for Richardson to strip the legal concept of bailment from his argument. He also refrains from calling researchers' obligations fiduciary. However, avoiding such concepts obscures the connections to legal obligations. Even though the analogy to bailment is not perfect, the underlying feature that spurs the law to recognize bailments is still in place, namely, even though a transaction may not be for the purpose of looking after someone or something's welfare, it may have the effect of making that individual or thing vulnerable to

another person's actions. In the case of medical research, a participant grants a researcher access to personal information to which she would not otherwise have access. Given the researcher's specialized knowledge, having such access sometimes puts her in a position of being able to benefit the participant in ways others cannot, as in the case of the heart infection.

It is not just that researchers are able to benefit their participants, though, that makes the participants vulnerable in the relevant respect. As we've seen, the vulnerability in fiduciary relationships stems from the transfer of power from entrustor to fiduciary, so it's important that researchers are able to benefit their participants as a result of the participants granting them permission to perform medical procedures. If a researcher just happened to notice a somewhat suspicious looking mole on a person's neck at the bus stop,³¹ he would not have a legal fiduciary duty to tell the person that she should have it checked out, since there is no basis for claiming that the person has empowered him to look after an aspect of her welfare. He may have a non-legally enforceable ethical obligation to tell him to get it checked out, but such obligations are not my focus here.³²

With this crucial feature of the researcher-participant relationship in view, we are now in position to see that it possesses each of the four features of fiduciary relationships, if only in attenuated form. First, we are justified in thinking that in entering a researcher-participant relationship, a participant gives a researcher (or research team) the power to look after aspects of his or her health. By giving permission to perform medical procedures and access private health information, participants open to the scrutiny of researchers broader areas of their health than are contained within the specific focus of the research study. Giving them this more generalized access is also to give them a more general power to look after these areas of their health.

Even though transferring this power is not the stated purpose of the relationship, we have seen that bailment relationships do not depend on the existence of any such stated purpose. But this feature is not restricted to bailment relationships. The existence of fiduciary relationships more generally depends primarily on the structure of the relationship, rather than the intentions of those who enter them. The law imposes fiduciary obligations on those who possess the characteristics of fiduciaries whether or not they, or their entrustors, intended to create a fiduciary relationship.³³ And by participants granting access to a range of health information and researchers accepting this access, researchers accept a role very similar to the physician fiduciary one. Given their specialized knowledge and their access to the participants' medical information, researchers, like physicians, are in position to provide participants with medical information that they would not be able to gain on their own. We ought to conclude that, as in the physician-patient relationship, researchers have been given the power to look after aspects of their participants' welfare. They may exercise this power by, for instance, reporting medically significant incidental findings back to participants.

Second, researchers possess some measure of discretion in deciding how to act in a participant's best interests. As the issue of incidental findings illustrates, researchers cannot always predict what medically relevant information they will uncover about a participant in the course of a trial. This is particularly true with research that is conducted on samples stored in biobanks. As technology has advanced, researchers are able to make discoveries in samples that were not possible at the time the samples were collected. Consider the infamous example of HeLa cells. When cancerous cells were taken from Henrietta Lacks in 1951, researchers could not have known all the purposes for which her cells would be used. Even if they had received informed consent from Ms. Lacks to use her cells in research, as they should have, they could have not indicated the range of possible uses in an informed consent form.

Third, researcher-participant relationships are founded on a mutually voluntary transaction, namely, an agreement to enter the research study. While the purpose of this agreement is not to entrust the participant's health to the researcher, we have seen that fiduciary relationships can exist even when a relationship is entered into for other reasons. The extent of a fiduciary's obligation of care will be more limited, however, when the relationship is entered into for other purposes since the fiduciary's primary responsibility remains fulfilling that purpose. For research relationships, this means that the objective of producing generalizable knowledge can limit the extent of the obligation to benefit participants.

Finally, researchers possess some degree of responsibility to act in the interests of their participants. As I have mentioned, this responsibility, like the responsibility in full-fledged fiduciary relationships, is based on the vulnerability that the research-participant relationship introduces. This has an important implication for research relationships. Informed consent is generally treated as the main legally and ethically important feature of the research relationship. And this is as it should be. Receiving informed consent from participants respects their right to make autonomous decisions about their participation in research. However, it is not the only legally and ethically important feature of research. For instance, researchers must not expose participants to excessive risk even when participants are fully willing to take on such risk. By holding that the researcher-participant relationship is partially fiduciary, we can better see why informed consent is not all that matters in research. Paying attention to participants' vulnerability shows us that researchers do not shed their responsibility to consider a participant's interests by handing potential participants an informed consent form.³⁴

Therefore, although the researcher-participant relationship departs in some respects from standard fiduciary relationships, it also possesses elements of their main features: transfer of power, discretion, a mutually voluntary agreement, and an obligation to act in the entrustor's interest.

4. BIOBANKS

Before wrapping up, it's worth noting that in studies on biological samples, an additional departure from standard

fiduciary relationships exists. The law has been uniform in its treatment of biological samples as donations.³⁵ For example, in *Washington University v. Catalona*, the Eastern District Court of Missouri held that Washington University retained ownership of biological samples when a faculty researcher, William Catalona, departed for Northwestern University, even though eight people who donated samples expressed a desire for their samples to go with Catalona to Northwestern.³⁶ (Catalona had sent a letter to those who had donated samples, asking that they make such requests, so that he could bring the store of samples with him to Northwestern.) As donations, ownership of the samples passes over to the party to whom the donation is given.

Fiduciary relationships, however, are based around the idea that ownership is not transferred during the relationship. Fiduciaries look after something that is not their own. To use my earlier example, a fiduciary relationship is not established if I give you my prized baseball card collection and ask you to take care good care of it. But one is established if I ask you to hold on to my collection for safekeeping while I travel around the world.

Can a fiduciary relationship exist even without this central feature? As before, I think the answer is that the relationship between those who donate samples and those who receive them may not technically count as fiduciary, but many of the same normative features remain. While the judgment that biological samples are donations seems correct, research participants still retain an interest in the contents of their samples, most notably the genetic material they contain. Like participants who actively partake in trials, those who donate biological samples grant researchers access to private medical information. When performing genetic tests on biological samples, researchers are increasingly finding genetic information that is medically useful, such as mutations in the BRCA genes, which are known to increase risk of breast cancer. Granting access to this medical information, therefore, places those who donate samples in the same position of vulnerability as those who actively participate in trials. And so the same rationale for an obligation of care carries over.³⁷ The law therefore ought to recognize that while biological samples are gifts, they ought to be treated differently than other sorts of gifts. Researchers working with such samples carry some measure of responsibility to act in the participants' interests, such as reporting incidental findings back to the original researcher, who can then re-contact the participant.³⁸ (Researchers working with samples from biobanks must rely on the original researchers to relay such findings since samples in biobanks have their identifying features removed.³⁹) Recognizing such a responsibility is critically important since as testing improves and its cost drops—whole genome sequencing now costs just a few thousand dollars—such findings will continue to be more common.⁴⁰

5. CONCLUSION

Despite this importance, the law has not, in any case of which I am aware, actually recognized that researchers bear fiduciary obligations to their participants. Indeed, in at least one case, *Hamlet v. Genentech Inc., et al.*, a court has

ruled that researchers do not possess fiduciary obligations. In this case, the North Carolina Superior Court ruled that the defendants, including the researcher, "are not health care providers" and "do not owe a fiduciary duty to Plaintiff as a result of Plaintiff's participation in the study."⁴¹

It is understandable that courts would not recognize researchers as fiduciaries. As I have noted, they differ in some important ways from standard fiduciaries. Most notably, their primary objective is to generate generalizable knowledge, not to benefit their participants. Those who argue that researchers bear all the main characteristics of fiduciaries, such as Paul B. Miller and Charles Weijer,⁴² therefore overreach. Instead, as I have argued, while researchers are not full-fledged fiduciaries, the type of vulnerability present in researcher-participant relationships is the same as that in legally recognized fiduciary relationships. Since this vulnerability grounds obligations of care in standard fiduciary relationships, it does the same in researcher-participant relationships. The law therefore ought to recognize that while researchers are not full-fledged fiduciaries, they do possess at least an attenuated form of the same obligation as fiduciaries. The Maryland Court of Appeals' ruling in *Grimes v. Kennedy Krieger Institute* provides an example of the way ahead. There, the court ruled that a "special relationship" exists between researchers and participants, which gives researchers the responsibility to look after their participants' welfare, but refrained from describing this relationship as fiduciary.⁴³ This example shows that the law needn't classify researchers as technically fiduciaries to acknowledge that they shoulder an obligation to benefit their participants. While the failure to fall neatly into any pre-established type of legal relationship may make legal treatment of the researcher-participant relationship difficult, the law ought to recognize the normatively important particularities of the relationship. By resisting the temptation to force the relationship into a pre-established category, the law can acknowledge participants' vulnerabilities while still distinguishing researchers from clinicians.⁴⁴

NOTES

1. AMA's Code of Medical Ethics, available at <http://www.ama-assn.org/ama/pub/physician-resources/medical-ethics/code-medical-ethics.page>, accessed April 15, 2014. For legal opinions, see, e.g., *Moore v. Regents of the University of California*, 51 Cal. 3d 120; 271 Cal. Rptr. 146; 793 P.2d 479 (1990) and *Hammonds v. Aetna Casualty & Surety Company*, 243 F.Supp. 793 (N.D. Ohio, 1965).
2. See, e.g., *The Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research*, National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, 1979, <http://www.hhs.gov/ohrp/humansubjects/guidance/belmont.html>, accessed June 18, 2014.
3. Susan M. Wolf, et al., "Managing Incidental Findings in Human Subjects Research: Analysis and Recommendations," *Journal of Law, Medicine, and Ethics* 36 (2008): 226.
4. E.g., Wolf et al., "Managing Incidental Findings," "Anticipate and Communicate: Ethical Management of Incidental and Secondary Findings in the Clinical, Research, and Direct-to-Consumer Contexts," 2013, <http://bioethics.gov/node/3183>, accessed April 15, 2014; and Henry S. Richardson and Leah Belsky, "The Ancillary-Care Responsibilities of Medical Researchers," *The Hastings Center Report* 34, no. 1 (2004): 25–33.
5. Henry S. Richardson and Mildred K. Cho, "Secondary Researchers' Duties to Return Incidental Findings and Individual Research

- Results: A Partial-Entrustment Account," *Genetics in Medicine* 14 (2012): 467-72.
6. J. C. Shepherd, *The Law of Fiduciaries*, Toronto: The Carswell Company Limited, 1981; as cited in Haavi E. Morreim, "The Clinical Investigator as Fiduciary: Discarding a Misguided Idea," *Journal of Law, Medicine, and Ethics* 33:3 (2005): 588. See also Tamar Frankel, "Fiduciary Law," *California Law Review* 71:3 (1983): 795-796.
 7. Paul B. Miller and Charles Weijer, "Fiduciary Obligation in Clinical Research," *Journal of Law, Medicine, and Ethics* 34:2 (2006): 425; and Deborah A. DeMott, "Beyond Metaphor: An Analysis of Fiduciary Obligation," *Duke Law Journal* 879 (1988): 879.
 8. In treating inequality and the potential for abuse of power as defining features of fiduciary relationships, my analysis follows Frankel's. See Frankel, "Fiduciary Law," 808-816.
 9. The law holds that entrustors in general are entitled to trust fiduciaries (Frankel, "Fiduciary Law," 824).
 10. The inequality I describe in this section is restricted to the item or aspect of one's welfare that is entrusted. An entrustor may, all things considered, actually possess more power than her fiduciary. For instance, Warren Buffett's physician has the power to look after his health, but has much less power than Buffett overall, I presume.
 11. Technically, in a legal trust, I pass legal title to you but retain "beneficial ownership." That is, I pass the power to you, but retain the benefits of ownership. For example, if you sold the baseball cards while I was gone, I would still be entitled to the value of the cards. (See Frankel, "Fiduciary Law," 828.) So, by "retaining ownership," I mean retaining beneficial ownership.
 12. See Henry Richardson's discussion of "forward-looking moral responsibilities" for a fuller discussion of the need to account for unexpected circumstances shapes the character of responsibilities (Henry S. Richardson, "Institutionally Divided Moral Responsibility," *Social Philosophy & Policy* 16 (1999): 218-49).
 13. One exception to this criterion is the parent-child relationship. In at least some jurisdictions, the law has recognized parent-child relationships as being fiduciary in nature, but children do not choose to entrust their parents with their care. It is also arguable whether parents become fiduciaries through voluntarily accepting parental obligations. Some accounts of parental obligations treat them as voluntarily assumed, but others do not.
 14. E.g., Miller and Weijer, "Fiduciary Obligation in Clinical Research," 431-432.
 15. *Meinhard v. Salmon*, 249 N.Y. 458 (N.Y. 1928), available at http://www.nycourts.gov/reporter/archives/meinhard_salmon.htm (accessed June 10, 2014).
 16. See Frankel, "Fiduciary Law," 820-824 and DeMott, "Beyond Metaphor," 922-23.
 17. See Robert E. Goodin, *Protecting the Vulnerable* (Chicago: The University of Chicago Press, 1985): 110. Goodin does not separate mistreatment from harm, but I do in order to accommodate views for which exploitation need not involve harm.
 18. Morreim, "The Clinical Investigator as Fiduciary," 588-589; Miller and Weijer, "Fiduciary Obligation in Clinical Research," 431.
 19. Miller and Weijer, "Fiduciary Obligation in Clinical Research," 433.
 20. Other objections exist too, but they are less significant. Haavi Morreim, for instance, argues that researchers do not possess the discretion that fiduciaries do (Morreim, "The Clinical Investigator as Fiduciary," 589-590). Paul B. Miller and Charles Weijer, however, ably respond to this charge (Miller and Weijer, "Fiduciary Obligation in Clinical Research," 429-430).
 21. Morreim, "The Clinical Investigator as Fiduciary," 590.
 22. Neal Dickert and David Wendler, "Ancillary Care Obligations of Medical Researchers," *JAMA* 302:4 (2009): 424-8.
 23. MedlinePlus (US National Library of Medicine), accessed April 15, 2014, <http://www.nlm.nih.gov/medlineplus/ency/article/001098.htm>.
 24. Richardson and Belsky, "The Ancillary-Care Responsibilities of Medical Researchers," and Leah Belsky and Henry S. Richardson, "Medical researchers' ancillary clinical care responsibilities," *BMJ* 328 (2004): 1494-6.
 25. Henry S. Richardson, *Moral Entanglements: The Ancillary-Care Obligations of Medical Researchers* (Oxford: Oxford University Press, 2012).
 26. Richardson and Belsky, "The Ancillary-Care Responsibilities of Medical Researchers," 28. Richardson and Belsky suggest that bailment relationships are not a type of fiduciary relationship, but others disagree. See, e.g., Frankel, "Fiduciary Law," 795.
 27. Richardson and Belsky, "The Ancillary-Care Responsibilities of Medical Researchers," 28.
 28. *Ibid*, 28.
 29. Richardson, *Moral Entanglements*, 39n.
 30. See chapter 3 of Richardson, *Moral Entanglements*.
 31. This example comes from Richardson, *Moral Entanglements*, 76.
 32. There is a legal duty to warn of foreseeable harm, and some appeal to this duty to ground the duty to report incidental findings (e.g., Wolf et al., "Managing Incidental Findings," 229-230). I appeal to fiduciary duties because they ground obligations in a broader range of cases. For example, a physician needn't tell the person at the bus stop about the mole, but he should tell his own patient. Likewise, it seems that researchers ought to tell their participants about a broader range of incidental findings than are covered by the legal duty to warn.
 33. Frankel, "Fiduciary Law," 820-821.
 34. My argument here and my reference to the Lacks case are not intended to oppose efforts to improve informed consent forms to give participants more say in how their samples are used, such as through tiered consent. It does imply, however, that such efforts can be only partial solutions.
 35. See, e.g., *Moore v. Regents of the University of California*, *Greenberg v. Miami Children's Hospital Research Institute* 264 F.Supp.2d 1064 (S.D. FL, 2003), and *Washington University v. Catalona* 437 F.Supp.2d 985 (E.D. MO 2006).
 36. *Washington University v. Catalona*.
 37. See Richardson and Cho, "Secondary researchers'," for an argument that the specific rationale for the partial-entrustment model carries over to researchers working on biological samples.
 38. Richardson and Cho, "Secondary researchers'."
 39. *Ibid*.
 40. Others have argued that those who donate biological samples have an autonomy right to control the use of their genetic material. My argument in this section is not intended to oppose such arguments. Rather, it can complement them. Thanks to Stephanie Alessi for raising this point.
 41. As quoted in Morreim, "The Clinical Investigator as Fiduciary," 592.
 42. Miller and Weijer, "Fiduciary Obligation in Clinical Research."
 43. *Grimes v. Kennedy Krieger Institute* 36 Md 29, 782 A2d 897 (2001), available at <http://www.courts.state.md.us/opinions/coa/2001/128a00.pdf> (accessed June 10, 2014).
 44. Previous versions of this paper were presented at the Northern Arizona University Philosophy of Law Conference and a Stanford University Center for Integration of Research on Genetics and Ethics writing seminar. I received many helpful comments from the attendees at both the conference and the writing seminar. Thanks in particular to Mark Budolfson, Luke Maring, and Katharine Schweitzer, my co-presenters at the conference; Christopher Griffin and Steven Scalet, the conference organizers and editors of this journal; and Stephanie Alessi and Mildred Cho, the commentators at the writing seminar.

Inquiring into Nature of the Legal Good

Katharine Schweitzer
UNIVERSITY OF NEVADA, RENO

Robin West, the Frederick J. Haas Professor of Law and Philosophy at Georgetown Law, exhorts legal scholars to turn their attention to the nature of the “legal good,” that is, to “the quality or qualities of goodness that a good law possesses.”¹ West has taken stock of the current state of legal scholarship, and she finds it wanting. I quote at length from her diagnosis:

With only a few exceptions, the major practitioners of our dominant contemporary jurisprudential movements do not ask what justice requires of law, what makes a good law good or what makes a bad law bad, or what the good is that law can, or should, accomplish, against which we might judge particular laws or legal regimes. We do not ask how we know that a good law is a good or a bad law, or whether those assumptions are warranted or unwarranted. Even more clearly, we do not often ask what it is about social life that seemingly requires address by or recompense from the law, or how a good law might respond to a social ill in a desirable fashion, or how a bad law, or no law, might do so poorly. We do not often ask whether we have insufficient law to address ills of private or social life, or how we would know that, if it is true. We do not often ask, anymore, whether the legal regime in which we live and which we have constructed with law is promoting good lives for us and our neighbors.²

In sum, according to West, most legal theorists have ceased to engage in scholarship that is normative in nature. Current practitioners of the three most prominent jurisprudential movements—natural law, legal positivism, and critical legal theory—do not inquire into how and why law contributes to individual and collective human flourishing. Although many legal scholars theorize about the procedural fairness of legal processes and about “whether a law maximizes wealth, efficiency, or benefits over cost,” West argues that very little legal scholarship considers the relation of law to justice.³

I take West’s criticisms of the scholarship produced by members of law faculties as a starting point for examining the division of cognitive labor in theorizing normatively about law. Epistemologists and philosophers of science introduced the term “the division of cognitive labor,” which they use to describe the social structures and processes through which knowledge about an issue is pursued and is assessed.⁴ In these subfields, philosophers “have become more sensitive to the social nature of scientific enquiry.”⁵ They have produced different models of how cognitive labor can be divided among individual scientists and how scientists might choose their projects and then communicate the results of their research to other scientists.

Similar questions about the division of cognitive labor should be asked about the production of knowledge of law, specifically what social problems law aims to solve and how law contributes to just political governance and human flourishing. In two recent monographs that defend inquiry into the nature of the legal good, West directs her arguments to professors who engage in law school teaching and scholarship.⁶ She claims that law professors ought to pursue questions that political theorists, moral philosophers, and scholars of public policy currently ask. She does not describe normative jurisprudence as a project which political theorists, moral philosophers, and scholars of public policy might undertake with legal scholars. After adumbrating West’s account of normative jurisprudence, my first aim in this paper is to theorize about how philosophers who work in the area of value theory ought to proceed if they wish to contribute to inquiry into the nature of the legal good and to advocate for what law ought to be. I draw on West’s earlier scholarship in developing my claims about the value of including philosophers in the conversation about what good law ought to do.

My second aim is to improve on West’s arguments that normative jurisprudence is a project that is worth pursuing. I accomplish this aim by first theorizing about how to resolve disagreements between competing conceptions of the legal good and then by offering a pragmatic theory of the nature of the legal good. West’s account of how scholars ought to understand “the justice law ideally serves” includes many features that I identify as embodying the normative commitments of some theories of philosophical pragmatism.⁷ In the second half of my paper, I defend this interpretation and argue that the project or normative jurisprudence that West encourages scholars to take up will be more successful if West’s account of this endeavor becomes more self-consciously pragmatic in methodology.

West’s legal scholarship is not well-known among philosophers. A search of her name in the *Philosopher’s Index* produces no articles that refer to her work. Some philosophers who work in feminist theory may know West on account of Janet Halley’s forcefully critical review⁸ of West’s 1997 monograph *Caring for Justice* and West’s equally critical response.⁹ Martha Nussbaum has also published, albeit in a law review,¹⁰ on West’s 1988 article “Jurisprudence and Gender,” a piece that made an important contribution to the development of feminist legal theory.¹¹ By explicating some of West’s theses and putting these claims in relation to contemporary debates within philosophy of law about the relation of law to morality and about legal pragmatism, I hope to generate interest among more philosophers in her work.

INQUIRING INTO THE NATURE OF THE LEGAL GOOD

Contemporary jurists, according to West, look into the relationship of law and positive morality and into the source of law’s authority. Legal scholars, however, do not theorize about whether and how law and legal institutions can and should play a role in solving social problems and in creating a just society. West argues against this status quo. Justice is the end of which law is the means, she

claims. On her view, the purpose of law is to secure justice. Consequently, justice should be a topic of legal scholarship and arguments about the correct interpretation of legal doctrines ought to appeal to a substantive theory of justice.

West's inquiry into the nature of the legal good stands in stark contrast to contemporary jurisprudential inquiry into the relation of law and morality. Legal positivists attempt to understand the conditions of legal validity and to articulate criteria for what counts as law. Legal positivists who engage in analytic jurisprudence do not understand their scholarship as having a relation to justice and to criticizing whatever it is that counts as law as against an account of justice. In contrast, West inquires into the value of the practice of regulating our lives according to law. West distinguishes the activity of "specifying the necessary and sufficient conditions for the existence of jurisprudentially significant phenomena" from the activity of inquiring into what a particular jurisprudentially significant phenomenon allows human beings who live in accordance with political rules to do and to be.¹²

West notes correctly that political philosophers, on the other hand, pay considerable attention to justice. Theories of political justice are generally substantive and provide an account of the nature and ends of political association. A theory of justice describes the proper relation between citizens and the state and how citizens ought to relate to each other. According to West, a theory of legal justice ought to build on a theory of political justice by explaining the role of law and legal institutions in facilitating proper relationships between political actors. A theory of legal justice offers an account of how law can be a tool for achieving justice.

The scholarship of Ronald Dworkin may initially appear to be a counterexample that undermines West's claim that contemporary legal philosophers lack interest in normative jurisprudence. Dworkin famously argued in *Taking Rights Seriously* that the law includes moral principles as well as legal rules and decisions and that legal argument is a morally inflected activity. West finds value in Dworkin's antipositivist approach because he thought seriously about the purpose of law. West is critical, however, of the results of his inquiry. His theory of law as integrity identifies justice with fidelity to the past. West argues that Dworkin's account of law as integrity can produce conclusions that offend our ordinary intuitions about justice. Even if we understand our current legal practice and the history of our legal tradition to include certain moral principles, these resources may not be sufficient for securing justice. We may need a moral commitment that is radically new and for which we cannot account using past interpretations of legal texts and decisions.

Moreover, according to West, Dworkin grounds his prescriptive claims about legal argument in descriptive claims about legal practice that many legal practitioners have contested. He would have done better to state explicitly why we should commit ourselves to particular moral principles instead of justifying these moral principles about what justice means and requires by claiming that they follow necessarily from a correct understanding of our legal

tradition. West laments that few legal scholars have been inspired to improve on Dworkin's account of the nature of the legal good and of how legal actors, specifically judges, ought to interpret law in light of moral principles.

According to West, Martha Nussbaum has articulated a theory of political justice that has promise as a compelling account of the nature of the legal good. The central thesis of Nussbaum's capabilities approach is that a just state must ensure that all citizens have the opportunity to enjoy ten central capabilities, which she understands as ways of being able to do things and to be in the world. According to Nussbaum, people who are denied the meaningful opportunity to develop these capabilities are also denied the prerequisites of a life that is worthy of human dignity.¹³ She envisions states codifying their commitment to providing citizens with the central capabilities in their constitutions. West considers Nussbaum's account of the ten capabilities as a promising start to a theory of the legal good. Where Nussbaum's theory falls short, in West's view, is in her failure to make a legal argument that existing constitutions provide legal actors with the authority or the obligation to change the legal status quo. Nussbaum's theory of justice and her account of the role of constitutional law in promoting her substantive account of justice might persuade political philosophers who are comfortable with working in the domain of ideal theory, but it is unlikely to persuade legal scholars who expect normative arguments about law to be grounded in authoritative legal sources and in doctrinal analysis. West maintains that Nussbaum's capabilities approach currently lacks an account of how states can find a commitment to protecting citizens' capabilities in their existing constitutions or a *legal* argument rather than a mere *moral* argument that states ought to revise their constitutions to include such a commitment.¹⁴ On her view, moral philosophers could articulate more effective social criticism if they could identify and incorporate existing legal resources into their scholarship.

THE DIVISION OF COGNITIVE LABOR IN THEORIZING ABOUT THE LEGAL GOOD

In the previous section I presented two prominent attempts by philosophers to theorize about what good law ought to do. Dworkin attempted to include resources from moral and political philosophy in his account of legal justice as fidelity to the best interpretation of our legal tradition. Nussbaum attempted to include resources from legal theory in her account of the constitutional obligations to citizens that states must fulfill in order to be classified as just. West assesses both approaches to normative jurisprudence as ultimately lacking. I maintain that we can draw an important conclusion about the desirability of cross-disciplinary efforts from an examination of these attempts to theorize about the legal good and West's commentary. Legal scholars and philosophers should work together to develop an account of what good law ought to promote that reflects the distinct and respective argumentative skills and knowledge that are characteristic of law and of value theory.

Inquiry into the nature of the legal good ought to fall within the purview of both academic disciplines. Philosophers who theorize about political morality ought to draw on

theories of constitutionalism and doctrinal analysis in order to locate textual support for the constitutional obligation to bring about the philosopher's preferred theory of substantive political justice. West suggests that engaging with constitutional theory will enable philosophers to appeal to recognized sources of authority, which will increase the motivation of legal actors who are bound by legal authority to consider normative accounts of what this authority demands. Given the way philosophers are currently trained in the United States, many political philosophers are not likely to have not been trained to understand the meaning and the content of particular bodies of law and in how to make arguments that persuade judges and lawmakers. West argues that value theorists ought to work to acquire the skill of doctrinal analysis and to test their "moral claims about justice against the weight of legal or judicial authority."¹⁵

Legal scholars who carry out doctrinal analysis of existing law and who theorize about the meaning of constitutions should draw on theories of the purpose and nature of political association. West suggests that engaging with value theory will enlarge the legal scholar's capability to envision what constitutional obligations a state might have beyond the current interpretation of what the courts have identified as constitutional obligations. Given the way law students are currently educated in the United States, many legal scholars are not likely to have received training in how to make sense of "raw appeals to intuition, moral sentiment, or vaguely felt and rarely articulated deontic moral principles" and to use these resources to develop philosophically sophisticated accounts of principles of substantive justice by which to assess the law as it currently exists.¹⁶ West suggests many internal reforms to teaching law and to legal scholarship. In her view, law students ought to learn about different accounts of what justice means and requires. Lawyers, judges, and lawmakers ought to understand justice as "the central point" of what they do.¹⁷ Law professors ought to understand justice as the ideal by which they should evaluate the merits of laws. West recommends that legal actors both study theories of justice that philosophers have articulated and improve on these theories by developing original accounts of what makes law or a political order just.

In addition to gaining familiarity with the different strategies of argument and the bodies of knowledge that are particular to the disciplines of law and value theory, scholars from both fields who engage in normative theorizing about the law ought to make an effort to share their research in joint communities of inquiry, such as workshops, conferences, and journals. Participating in cross-disciplinary exchanges can enrich the inquiry into the nature of the legal good in at least two ways. First, participants would be challenged to articulate and develop arguments in ways that scholars from other fields can understand and evaluate for truth and relevance. Second, participants would raise methodological questions about how to generate knowledge of the good that they wish law to accomplish.

Talking to other scholars, however, is not sufficient if a goal of theorizing about the legal good is to modify intelligently our existing political and legal practice

such that our shared common life might become more just. Consequently, West argues that theorists ought to "revitaliz[e] our constitutional culture" by engaging with the legislative branch of government at the state and federal levels and with lobbyists and social-movement actors who attempt to persuade legislators to use law as a tool for responding to and remedying social problems.¹⁸ Attending to traditions of constitutional interpretation outside of the courts, in both legislatures and in public political life, is an underappreciated context in which to think collectively about what kind of polity we wish our constitutional and statutory law to make possible. Engaging in "the ordinary, intensely political work of deliberative democracy" within public settings and legislatures often provides opportunities for citizens who hold competing accounts of substantive justice to persuade their opponents to correct their initial views or to construct compromises and coalitions so that we can pass ordinary laws to respond to social needs.¹⁹ If legal academics and philosophers become better at discussing amongst themselves the nature of the legal good, it may be more realistic to imagine that academics from a range of disciplines could address their arguments to the general public and engage with it. Both academic and nonacademic members of the political community stand to benefit from a more just social world. Scholars who wish to take up the project of normative jurisprudence should recognize that ordinary citizens ought to play a role in the cognitive division of labor in theorizing about the legal good.

WEST'S ACCOUNT OF THE LEGAL GOOD

West both theorizes about the need to inquire into the nature of the legal good and offers in her earlier work a substantive account of the goodness of good law that is grounded in a progressive conception of legal justice. West understands her account of the legal good as grounded in doctrinal analysis of the Constitution. She identifies the Fourteenth Amendment to the U.S. Constitution as textual authority for her progressive account of the legal good. The Fourteenth Amendment's Equal Protection Clause states that "No state shall [...] deny to any person within its jurisdiction the equal protection of the laws."²⁰ West removes the negative phrase from the constitutional text and argues that states have a positive obligation to protect their citizens and to protect them equally. She argues that the text of this Reconstruction amendment was intended originally to guarantee to the newly freed slaves the protection of criminal and civil law against lynching and other kinds of private violence that white citizens might commit.²¹ West explains,

Negatively, the clause requires the state to ensure that every individual is equally free of all conditions which could potentially subjugate his will to some sovereign power other than the state. The clause prohibits a state of dual sovereignty: the state must ensure that no citizen lives under more than one sovereign. The clause affirmatively requires the state to protect each individual's positive liberty: to guarantee to every individual the freedom to direct her own life and work.²²

West identifies the Equal Protection Clause as a textual basis for the existence of “unenumerated, positive, constitutional rights.”²³ By appealing to Section Five of the Fourteenth Amendment, which states that “the Congress shall have the power to enforce, by appropriate legislation, the provisions of this article,” West argues that both judges and legislators have an affirmative duty to provide the protection to which citizens have a constitutional right.²⁴ According to West, justice requires that law meets citizens’ social need for protection against certain kinds of harm. She argues that when the courts choose not to interpret the constitutional text in a way that recognizes this guarantee, state legislators have an obligation to propose laws that protect citizens and to do so equally and to attempt to pass these laws through the processes of democratic politics.²⁵

West’s understanding of the purpose of the state and the moral point of law, that is, of living according to rules, also draws on Thomas Hobbes’s social contract theory. According to Hobbes, people enter into political association because they desire protection that they cannot secure individually. In creating a social compact, the sovereign assumes the duty of protecting the people who relinquish their natural right to threaten or to punish others with the use of private force. West traces how the view that political power ought to be used to protect the welfare of the people because they have thus willed it can be found in other extralegal sources such as the writings of Thomas Paine.²⁶ West understands her progressive theory of justice as improving upon liberal theories of justice because her theory specifies how states should relate to citizens as beings who desire connection and intimacy with others as well as individuation and privacy from others. In contrast, liberal theories of justice understand human beings only as creatures who desire to exercise their wills and ignore important relational dimensions of human experience. West grounds her criticism of existing law in both an interpretation of what the existing law ought to be, that is, in an argument that legal actors consider as authoritative, as well as in moral ideals that people who are outside of the legal academy consider persuasive.

DEALING WITH DISAGREEMENT ABOUT THE NATURE OF THE LEGAL GOOD

Even though normative jurisprudence is not currently a rich field of inquiry, according to West, it is nevertheless possible to identify a plurality of accounts of the nature of the legal good. As some of West’s colleagues in the legal academy have noted in reviews of her work, many people disagree reasonably with progressive accounts of justice.²⁷ As I explained in the previous section, West’s understanding of legal justice includes an account of the affirmative duty of government to protect citizens from some kinds of harm. The state furthers legal justice when it responds to social problems with intelligent action. Dworkin articulated a liberal account of the legal good in which the role of good law is to protect the moral rights of individual people against government action. In order to facilitate justice, law must be used to constrain “a political community’s exercising its coercive power” to create policies that aim to promote the general welfare.²⁸ According to Dworkin, rights always trump considerations of the common good.

West defends the opposing view. How ought we to decide between competing accounts of a just relationship of human beings to law?

West does not pose and answer this question in the monographs in which she describes and defends normative jurisprudence. In her earlier work, West advocates a phenomenological jurisprudential method by which people describe the experience of being subject to a particular law or legal regime.²⁹ She considers it uncontroversial that laws that produce the experience of pain and suffering are *prima facie* bad and that laws that produce the experience of pleasure and subjective well-being are *prima facie* good. It seems to me that the method of describing one’s experience of a law or of the absence of a law in an attempt to persuade others to understand the legal status quo as good or bad is a promising way to reach consensus on the nature of the legal good. Such communication can be conducted in the context of open and inclusive public deliberation. Moreover, no single discipline can claim to be the authoritative arbiter of the terms of phenomenological description. An account of the nature of the legal good and a vision of the just society are successful when they garner popular support from the people whose lives are affected by the law or the absence of law. There are many ways by which a person’s description of the harm that law does or the harm that law ought to prevent can inspire sympathetic understanding in others.³⁰ West does not appeal to absolute, universal, and foundational truths about justice to defend her preferred substantive account of justice. Although West judges her arguments in support of her progressive understanding of the nature of the legal good to be superior to her rivals, she stresses that the success of an argument is revealed by its ability to persuade everyday citizens who reflect on their and others’ experience of law.

A PRAGMATIC THEORY OF THE NATURE OF THE LEGAL GOOD

Although West describes her substantive account of justice as progressive, I maintain that the public inquiry into the nature of the legal good that she recommends can be understood as pragmatic. Highlighting the pragmatic values implicit in West’s account of normative jurisprudence diffuses the worry that West takes the question of the nature of justice as settled conclusively rather than as open for argument and debate. “Legal pragmatism” is a disputed term and a wide range of approaches claims the title. My understanding of legal pragmatism draws on the philosophical pragmatism of John Dewey. In a 1985 article that bears the subtitle “A Pragmatic Definition of the Liberal Vision,” West articulates an interpretation of legal pragmatism that predates “the marked increase in legal academics’ interest in pragmatism” that occurred in the early 1990s.³¹ I argue that three features of legal pragmatism are salient in West’s recent work on normative jurisprudence. First, West understands law as a social tool by which to achieve intelligently a set of ends. Second, she employs a pragmatic theory of values that is rooted in first-person experience and does not appeal to abstract and universal moral principles. Third, West takes seriously the social, cultural, and historical contexts that inform lived experience. I briefly examine these features in turn.

A central claim in West's normative jurisprudence is that law is the means for achieving a desired end, namely, justice. West describes law as "the virtue to which this monumental shared societal work—the creation and running of a legal system—is aimed."³² Pragmatism is an approach that assesses the value of a given action by the practical consequences that action makes possible. Legal pragmatism that is informed by Dewey's philosophy stresses that people identify certain consequences as desirable ends because these consequences ameliorate perceived ills and promote perceived goods. West interprets Dewey's pragmatic liberalism as "at root committed to the attainment of the good life as the end of governmental action."³³ People identify ends through lived experience; they do not take the meaning of the good life as pre-given. According to West, Dewey argued that human beings must assess critically what consequences they take to be constitutive of the good life and to understand the end of human flourishing as "an evolving and social subject of study."³⁴ For West, understanding how citizens, legislators, and legal authorities can use law as a means to facilitate justice is an activity that first requires them to discuss and evaluate competing conceptions of justice and then requires them to use this understanding of justice to inform legal practice.

West also urges theorists who study the law to adopt a pragmatic theory of values and valuation. According to one account, to engage in philosophical pragmatism means to determine which views are "useful for some particular purpose in some particular situation."³⁵ Pragmatists understand value to be a relation between a person and his or her subjective experience. Because our subjective experiences are constantly changing, what we value also undergoes flux and transition. West's approach to theorizing about legal justice does not involve appealing to abstract and universal foundational moral principles. In contrast, West argues that we ought to acquire inductively the political values that a particular community holds at a particular historical moment. She explains, "A truly empirical method of value requires a cooperative, communitarian sharing of hypothesis, method, and tentative conclusion, just as empiricism in the natural sciences requires not elitism but cooperative inquiry by the relevant concerned scientific community."³⁶ West's approach to normative jurisprudence eschews the rational deduction of moral principles and instead focuses on first-person descriptions of the subjective experience of pain and pleasure. Accounts of justice should arise from lived experience, and people whose experiences differ from each other ought to examine in a spirit of cooperation the merits of competing accounts of how law might alleviate human suffering and facilitate human flourishing.

Last, West's account of legal justice is more sympathetic to the value of historical tradition than many liberal and radical accounts. Legal pragmatism values understanding and reflecting on past experience for the purpose of determining whether contingently held values merit continued support. According to West, pragmatism critically assesses inherited traditions and does not rule out from the start the possibility that traditional understandings of personal and communal life may be worth preserving and that traditional sources of

authority, including legal authority, have normative weight. Michael Sullivan, a theorist of legal pragmatism, argues that "pragmatic theories of valuation explicitly recognize their dependence on historical reflection in securing the realization of values in the changing circumstances of the present."³⁷

West might be more effective in persuading philosophers and other scholars to take up the project of normative jurisprudence by offering a pragmatic account of the value of inquiring into the nature of the legal good. Many philosophers of law may be content to analyze conceptually legal phenomena. At this time they may not be concerned with understanding justice or achieving justice. They may even be dissuaded from theorizing about justice because they anticipate how costly disagreement about the correct conception of justice will be. A pragmatic defense of the value of engaging in normative jurisprudence might convince skeptics that they ought to not be satisfied with conceptual and theoretical research efforts and that they ought to contribute their cognitive labor to the theoretical and practical problem of how to use law to intelligently facilitate human and communal flourishing.

CONCLUSION

I am sympathetic to Robin West's call to legal academics to engage in normative jurisprudence, and I wish to see scholars from a range of disciplinary locations contribute to conversations about what good law might do and how law can be used to bring about justice in historically situated political communities. In this paper I have presented West's account of what it might mean to pursue inquiry into the legal good and how participants to this endeavor ought to understand and arbitrate disagreements among competing accounts of justice, including her own progressive account. I have also drawn attention to some general features of legal pragmatism that are implicit in her descriptive of the program of normative jurisprudence. I suggest that West could strengthen her defense of the value of the inquiry into the nature of the legal good by articulating more explicitly the desirable social consequences that a diverse group of participants might bring about by theorizing together about how to use law to "work out solutions to social problems intelligently."³⁸

NOTES

1. Robin West, *Normative Jurisprudence: An Introduction* (Cambridge: Cambridge University Press, 2011), 3.
2. *Ibid.*, 2.
3. Robin West, *Teaching Law: Justice, Politics, and the Demands of Professionalism* (Cambridge: Cambridge University Press, 2014), 57.
4. See Philip Kitcher, "The Division of Cognitive Labor," *Journal of Philosophy* 87, no. 1 (1990): 5-22.
5. Ryan Muldoon, "Diversity and the Division of Cognitive Labor," *Philosophy Compass* 8, no. 2 (2013), 124.
6. Robin West, *Normative Jurisprudence and Teaching Law*.
7. Robin West, *Teaching Law*, 211.
8. Janet Halley, "The Politics of Injury: A Review of Robin West's *Caring for Justice*," *Unbound* 1 (2005): 65-92.
9. Robin West, "Desperately Seeking a Moralism," *Harvard Journal of Law and Gender* 29 (2006): 1-49.

10. Martha C. Nussbaum, "Robin West, *Jurisprudence and Gender: Defending a Radical Liberalism*," *The University of Chicago Law Review* 75, no. 3 (2008): 985-96.
11. Robin West, "Jurisprudence and Gender," *The University of Chicago Law Review* 55, no. 1 (1988): 1-72.
12. Matthew H. Kramer, "What is Legal Philosophy?" *Metaphilosophy* 43, nos. 1-2 (2012), 132.
13. See Martha Nussbaum, *Creating Capabilities: The Human Development Approach* (Cambridge, MA: Harvard University Press, 2011).
14. Robin West, "Human Capabilities and Human Authorities: A Comment on Martha Nussbaum's *Women and Human Development*," *Saint Thomas Law Review* 15 (2003): 763-90.
15. West, *Teaching Law*, 88.
16. *Ibid.*, 87.
17. *Ibid.*, 92.
18. Robin West, "Ennobling Politics," in *Law and Democracy in the Empire of Force*, eds. H. Jefferson Powell and James Boyd White (Ann Arbor, MI: The University of Michigan Press, 2009), 60.
19. Robin West, "Constitutional Culture or Ordinary Politics: A Reply to Reva Siegel," *California Law Review* 94 (2006), 1485.
20. U.S. Constitution, Fourteenth Amendment, Section 1.
21. Robin West, *Progressive Constitutionalism: Reconstructing the Fourteenth Amendment* (Durham, NC: Duke University Press), 23.
22. *Ibid.*, 31.
23. Robin West, "Unenumerated Duties," *University of Pennsylvania Journal of Constitutional Law* 9 (2006), 224.
24. U.S. Constitution, Fourteenth Amendment, Section Five.
25. Robin West, "Toward the Study of the Legislated Constitution," *Ohio State Law Journal* 72 (2011), 1360.
26. Robin West, "Tom Paine's Constitution," *Virginia Law Review* 89 (2003), 1416.
27. See Robert F. Nagel, "Progress and Constitutionalism," *Michigan Law Review* 94, no. 6 (1996): 1495-509 and Suzanna Sherry, "Progressive Regression," *Stanford Law Review* 47, no. 5 (1995): 1097-115.
28. Ronald Dworkin, *Justice for Hedgehogs* (Cambridge, MA: Harvard University Press, 2011), 329.
29. Robin West, "Jurisprudence and Gender," 64-65.
30. See West's description of sympathetic knowledge in "Disciplines, Subjectivity, and Law" in *Narrative, Authority, and Law* (Ann Arbor, MI: The University of Michigan Press, 1993), 292-8.
31. Michael Sullivan and Daniel J. Solove, "Can Pragmatism Be Radical? Richard Posner and Legal Pragmatism," *Yale Law Journal* 113 (2003), 687.
32. Robin West, *Teaching Law*, 60.
33. Robin West, "Liberalism Rediscovered: A Pragmatic Definition of the Liberal Vision," *University of Pittsburgh Law Review* 46 (1985), 694.
34. *Ibid.*, 695.
35. John J. Stuhr, "Pragmatism about Values and the Valuable: Commentary on 'A Pragmatic View on Values in Economics,'" *Journal of Economic Methodology* 10, no. 2 (2003), 214.
36. Robin West, "Liberalism Rediscovered: A Pragmatic Definition of the Liberal Vision," *University of Pittsburgh Law Review* 46 (1985), 698.
37. Michael Sullivan, *Legal Pragmatism: Community, Rights, and Democracy* (Bloomington, IN: Indiana University Press, 2007), 119.
38. Robin West, "Liberalism Rediscovered: A Pragmatic Definition of the Liberal Vision," *University of Pittsburgh Law Review* 46 (1985), 677.

The Truth in Public Reason: Why Retributivism Cannot Make Legal Punishment Legitimate

Luke Maring

NORTHERN ARIZONA UNIVERSITY

Either he [the retributivist] appeals to something else—some good end—that is accomplished by the practice of punishment, in which case he is open to the criticism that he has a nonretributivist, consequentialist justification for the practice of punishment. Or his justification does not appeal to something else, in which case it is open to the criticism that it is circular and futile.

– Hugo Bedeau, "Retribution and the Theory of Punishment," *Journal of Philosophy* (1978)

1. INTRODUCTION

Exactly *why* is it important that criminals suffer legal punishment? Bedeau claims that a retributivist's answer will fall into one of two traps. It will invoke consequentialist values or it will be circular. As stated, however, Bedeau's dilemma cannot be correct. There is too much space between its horns: Bedeau fails to explain why a retributivist cannot explain the moral worth of punishment by citing deontological values.

But Bedeau was close to correct. A dilemma very much like his, when propped up by the right theoretical machinery, shows that retributivism cannot make harsh legal punishments like imprisonment or execution legitimate. My argument begins with three clarificatory steps. First, I will explain what legal punishment is; second, I will define retributivism; and third, I will determine what is required to make punishments like imprisonment or execution legitimate. With the stage thus set, I will introduce my variation on Bedeau's dilemma.

2. WHAT IS LEGAL PUNISHMENT?

Retributivism is a view about what justifies legal punishment; the definition of legal punishment is therefore our first order of business. There is a substantial literature devoted to articulating legal punishment's necessary and sufficient conditions.¹ Rather than joining the fray, I want to achieve clarity about legal punishment by comparing the practice to something else. Restorative justice, like legal punishment, is a practice that aims to solve the problems posed by crime. By juxtaposing the two, I will pinpoint the elements of legal punishment that raise moral concerns.

Take a pair of examples as a starting point: Suppose that someone steals \$1,000 from me. I apologize for the bare appeal to intuition, but if the court's whole sentence is that the thief should return my \$1,000, it does not seem that she was punished. Thieves are generally not entitled to their ill-gotten gains, so it seems a mistake to construe simple repayment as punishment. Or suppose that an energy company makes a fortune by illegally devastating the environment in a particular region. If the court's sentence

is merely that the company must use its illegally obtained profits to clean up its own mess, I think it odd to say that the company is punished. But *why* do these sentences not count as punishment?

The answer lies in a distinction often emphasized by restorative justice: the distinction between punishing and requiring restoration. Crudely put, the aim of restoration is to reinstate prior conditions. Health can be restored after sickness; money can be restored after a theft; and a relationship can (usually) be restored after a fight. In the case of criminal wrongdoing, the restorative process must accomplish two things: first, it has to restore material conditions; second, it has to undo the normative damage that crime inflicts upon our relationships (even when the relevant relationship is as thin as co-citizen). At the conceptual level, there is no mystery about how to restore material conditions—one transfers resources from the offender to the victim. As a practical matter, it might be unclear just how much we should transfer. And there will be cases when a perfect material restoration is impossible (e.g., a thief takes and destroys something that is one-of-a-kind). My point, here, is that there is no conceptual puzzle about how it is possible to restore material loss, however imperfectly we manage to do it in practice.

But how can we repair the normative damage that a crime does to our relationships? Restorative justice usually relies upon a process of mediation in which, if all goes well, the offender is brought to (a) recognize her actions as constituting an offense; (b) apologize; and (c) undertake appropriate reconciliatory measures. The appropriate form of reconciliation will vary from case to case—it might involve a particular kind of community service, a certain gift to the victim, donations to a relevant charity, or, perhaps, counseling for an addiction or anger management problem. What matters is that those who are party to the crime understand the reconciliation as a demonstration of the criminal's sincere remorse.

How does restoration differ from punishment? Resorting once again to a crude slogan, whereas restoration aims to reinstate prior conditions (both material and normative), punishment aims to make criminals suffer. Different theorists cash out the suffering that constitutes punishment differently. Some characterize punishment as *pain*.² Others say punishment is a *harm*.³ Still others hold that punishment is a sort of *deprivation* or *loss*.⁴ For my purposes, the precise concept we use to analyze punishment does not matter. Requiring restoration is one thing; making a criminal suffer for the sake of inflicting pain, harm, deprivation, or loss is something else.

With the distinction between punishing and requiring restoration firmly in hand, return to the pair of examples I began with. The thief who is ordered to return my \$1,000 is not punished. The reason why should now be clear: the court chooses its sentence with an eye to reinstating prior material conditions, not to making her suffer. Similarly, the court's decision that the energy company must use its ill-gotten gains to clean up the ecosystems it has damaged is restorative rather than punitive.

Yet Anthony Duff has recently denied the distinction between punishing and requiring restoration by defending a strong modal claim: "the kind of restoration that criminal wrongdoing makes necessary is properly achieved through a process of retributive punishment."⁵ Initially, Duff's claim seems implausible. Restoration emphasizes mediation, apology, and reparation; punishment emphasizes suffering for suffering's sake. Why, then, does Duff think that restorative justice *necessarily* involves punishment?

[The offender] should be pained by the recognition of his wrongdoing . . . [these] pains are not mere side-effects of the process which—if he is lucky—might not ensue; they are integral to the aims of the process as a process of seeking restoration after a crime. That is to say, however, that the process of what we could call criminal mediation . . . is a *punitive* process.⁶

Duff's argument goes as follows: a genuine apology involves fully recognizing one's moral shortcomings; pain is a constitutive part of fully recognizing one's moral shortcomings; and restorative mediation aims at genuine apology. If we put these pieces together, we reach the conclusion that restoration necessarily aims to cause pain, and that, thinks Duff, makes restoration a punitive process.

I think there are counterexamples to Duff's reasoning. I recently read a referee report that laid bare the (substantial) flaws of a paper I've been working on. The referee intended that I would recognize my paper's shortcomings, and for me, at least, pain is a constitutive part of that process. Yet I think it odd to say that the sort of correction journal referees engage in counts as punishment, even for me. One might object that restorative justice aims for the recognition of a *moral* rather than an epistemic shortcoming, but even so there are counterexamples. Suppose that my partner catches me using sexist epithets and demands that I take an implicit bias test in order to lay bare the depth and breadth of my prejudice. Pain is, or ought to be, constitutive of genuinely recognizing one's moral deficiencies as one's own—the recognition that one, say, automatically regards women as inferior ought to hurt. But as with the journal referee, I think it odd to regard a moral diagnostic as punishment. So what's gone wrong with Duff's reasoning?

In a too-simple slogan, the answer is that whereas journal referees and moral diagnostics aim to educate, punishments aim for pain itself. The slogan is too simple because, as Duff points out, education can involve pain, even constitutively. But it is one thing to cause pain as part of a process that aims at something else; it is quite another to cause pain *for the sake of making the guilty suffer*.⁷ The difference, here, must be understood teleologically. The final end of restoration is to put things back as they were—and as Duff points out, that will for the offender necessarily involve the pain of recognizing her wrongdoing. But even though her pain is a necessary step along the way, the final end of restoration is not to cause suffering. The final end of punishment, by contrast, is precisely to make criminals suffer to the "appropriate" degree.

And *that* is why the practice of legal punishment raises moral concerns: it is *prima facie* wrong to cause suffering for suffering's sake. In the final tally, this *prima facie* wrongness might be outweighed. Perhaps we should make criminals suffer in order to satisfy the demands of justice. Or perhaps we should make criminals suffer because by doing so we will deter other would-be criminals. My point, in this section, is just to explain that legal punishment is a process whose final end is the infliction of suffering. Legal punishment therefore stands in need of justification.

3. WHAT IS RETRIBUTIVISM?

Legal punishment stands in need of justification; retributivism is a candidate justification. Unfortunately, with the advent of negative retributivism, the term "retributivism" has become a term of art. My aim in this section is two-fold. First, I will identify the retributive views that constitute my target. Second, I will use the distinction between punishing and requiring restoration to explain why retributivism is such a morally remarkable doctrine.

A comprehensive theory of legal punishment must answer "why," "who," and "how."

Why is it morally important to punish?⁸

Who should we punish?

How should we correlate punishments to crimes?

The "why" question is in two senses foundational. First, it asks why we should have a practice of legal punishment in the first place. "Who" and "how," by contrast, ask about administrative details that are important only if the practice of legal punishment is worth having. So if we cannot first answer "why," there is no reason to even entertain "who" and "how." (Compare: If there were no good reason to give cancer patients chemotherapy, an argument about who should get it and how would be beside the point.) Second, the answer to "why" will constrain answers to "who" and "how." An explanation of why legal punishment is morally important will, in a sense, say what the practice is good for. Our answers to "who" and "how," in turn, should be constructed with an eye to advancing, or at least to not undermining, the moral value that legal punishment is supposed to realize. For example, if deterrence is what makes the practice of punishment worthwhile, we should take care to choose punishments harsh enough to deter; if the practice of punishment aims to give criminals their just deserts, we should, by contrast, be especially concerned that the severity of punishment match the severity of crime.

Retributivism began as a comprehensive theory of legal punishment. It answered the foundational question of why: criminals' guilt calls for punishment. It answered who: those guilty of breaking the law. And most retributivists agreed about how: we should punish according to the principle of *lex talionis*, inflicting a suffering upon criminals equivalent to the suffering they inflicted upon their victims. *Lex* has largely fallen out of favor because of its barbaric implications—e.g., that the state ought to sentence rapists to rape. Fortunately for retributivism, *lex* is but one method for correlating punishments and crimes, and there is no

a priori reason that retributivists cannot answer "how" by citing a different principle.⁹

But while retributivism began as a comprehensive theory of punishment, one branch of retributivism has much lower ambitions. *Negative retributivism* is remarkable, in part, for how little it claims: guilt is a necessary, but not sufficient condition for punishment. Thus, negative retributivists have not only abandoned *lex*, they have abandoned retributivism's answer to the foundational question of "why." And, further, by insisting that guilt is merely necessary, they have substantially weakened the traditional retributivist answer to "who": whereas traditional retributivism holds that all the guilty ought to be punished, negative retributivism holds that some (possibly empty) subset of the guilty ought to be punished. These concessions were part of an attempt to form pluralist theories of punishment. My aim, here, is not to comment upon the viability of pluralist theories. It is to show that, in its weakened state, there is nothing retributive about negative retributivism.

Here is another way to put the point. Negative retributivism's *only* contribution is to identify guilt as a necessary condition of punishment. But "necessary" and "sufficient" are logical rather than justificatory relations. Thus, the slogan "punish only if guilty" does not say why guilt should precede punishment. Even a consequentialist could endorse "punish only if guilty" if that particular policy had the best overall result. (The consequentialist would probably say that certain social goods require citizens to regard their practice of legal punishment as legitimate, and that a practice risks its legitimacy by punishing the innocent.) Thus, to retain one's retributive *bona fides*, one must go beyond mere negative retributivism and take a stand on what makes punishment good in the first place, a stand on the foundational question of "why." In particular, a retributivist must insist that legal punishment is morally good, at least in part, *because it gives criminals what they deserve*.

The thought that punishment is good because it gives criminals what they deserve can sound like obvious commonsense. Kant, at least on some interpretations, suggests that retributivism simply follows from the concept of autonomy: we must hold criminals accountable to respect their autonomy, and we must punish them to hold them accountable. The first claim is arguably correct. In order to respect people as agents, we must recognize the decisions they make as theirs—holding them accountable when it is appropriate to do so. But even if we grant this first claim, the second claim is a *non sequitur*: it is one thing to insist that we must hold criminals accountable; it is quite another to insist that we must hold them accountable *by punishing them*. Restorative justice, after all, also aims to respect criminals' autonomy by holding them accountable. But holding accountable, for restorative justice, takes the form of requiring criminals to (a) acknowledge their wrongdoing and then to (b) undo, as far as possible, the material and normative damage they have wrought. Retributivists, by contrast, see guilt as not merely calling for apology and reparation, but as calling for punishment. *Pace* Kant, this is a substantive moral view and does not simply follow from the concept of autonomy. Retributivism needs a defense.

And a defense of retributivism must be powerful, as the view that guilt calls for punishment is morally remarkable. Permit me to drive home retributivism's strangeness: Someone commits a crime, and, as a result, acquires the moral status of guilty. But suppose that she recognizes her wrong, apologizes to the victim, and makes fully adequate reparations. We can even suppose that the victim genuinely forgives her and wants to let the matter rest. Retributivists might be glad to hear about these developments, but they think that, strictly speaking, the criminal's guilt has not been expunged—she is still guilty; there is still an excellent reason to punish her. Of course, a retributivist is not committed to saying that the reason to punish overrides all other reasons. Perhaps the appropriate punishment will take the criminal out of society for a long time and the downsides of that result may outweigh the reason to punish (suppose she has children and can, now that she is reformed, be expected to do great things for society). But even so, the retributivist is committed to a remarkable claim: it is in a sense morally regrettable that the reformed criminal escapes punishment; it would be better if we could somehow punish her without causing such bad consequences. But *why*, exactly, would it be a good thing to make such a person suffer? Why, in other words, does guilt call out for punishment in particular?

4. THE TRUTH IN PUBLIC REASON

The *telos* of punishment is suffering. What could make a practice of making criminals suffer legitimate? My first step towards an answer is to criticize public reason, though my aim is not decisive refutation. I will use my (not entirely original) criticism as a springboard to argue that the importance of what I will, following David Enoch, call justification-to varies from law to law.¹⁰

The literature on public reason has exploded over the past decade. So rather than taking any particular articulation of the idea as my starting point, I will proceed from a generic sketch of what public reason liberals have in common. At its core, the purpose of public reason is to provide the necessary and sufficient condition of political legitimacy.

A law is legitimate iff it can be justified to all the reasonable persons it will coerce in terms of reasons they can accept.¹¹

Notice that according to this criterion of legitimacy, a law is not justified in virtue of being the objectively best policy. The public reason liberal's thought is rather that a law's legitimacy hangs on its *justifiability-to*—it's being supported by reasons that the coerced would accept. David Estlund defends this claim with a thought experiment. He stipulates that Catholicism is the true religion and that the Pope, with his direct line to God, is infallible. The Pope's infallibility aside, Estlund continues, it would be wrong for the government to force papal edicts on non-Catholics.¹² The reason truth does not imply justification all on its own is that citizens are *free* and are the *moral equals* of the government officials who would coerce them. Citizens' freedom means that, *prima facie*, they should be allowed to choose their religion (or to choose atheism), even if they are certain to be mistaken. The moral equality between citizens and government officials means that government

officials have no automatic right to overrule citizens' freedom. It is important to see that nothing turns on the religious nature of Estlund's example. His point is the general one that it seems *prima facie* wrong to curtail the freedom of one's moral equals without first satisfying the demands of justification-to.

Thus far, the public reason liberal's view is abstract and, I think, compelling. Problems emerge when we try to specify the set of citizens to whom justification is owed. Actual citizens can be irrational, uninformed, or dogmatic, so it seems that *no* law could possibly win every actual citizen's approval. Does public reason's criterion of legitimacy therefore imply philosophical anarchism? To resist that conclusion, public reason liberals restrict the class of persons to whom justification is owed. We cannot justify laws to all actual citizens, but, their thought goes, we *can* justify laws to those who are reasonable. But I doubt the public reason liberal's restriction to the reasonable solves the problem. While Estlund's Pope shows that justifiability-to is a value for politics, it is but one value among many. Others include political stability, security, economic prosperity, fairness, public health, etc. There is a *diverse plurality* of values for politics and it is overwhelmingly plausible that reasonable citizens will rank them differently. Suppose, then, that a particular law *L* enjoys majority support and would advance several political values. But because some citizens overridingly prefer to spend their limited resources differently (in pursuit of other political values), *L* is not justifiable to every citizen. The public reason liberal's criterion thus implies that *L* is illegitimate. But notice: in calling *L* illegitimate, the public reason liberal gives justification-to absolute priority over the values *L* serves. In fact, because the values in this example are arbitrary placeholders, the public reason liberal gives justification-to absolute priority over every other political value. That is a bold moral claim. Why, after all, should justification-to trump every other political value on every occasion?

There are two sorts of answers prominent in the public reason literature. The first defends the priority of justification-to on substantive moral grounds. According to James Boettcher, for example, "respect for others as ends in themselves authorizes the liberal insistence that some form of reasonable agreement is what establishes the legitimacy of coercive political principles."¹³ But wait: "respect for others as ends in themselves" is, obviously, a Kantian formulation, and Kantianism is precisely the sort of sectarian view that public reason was supposed to do without: the public reason liberal condemns religious justifications because not all reasonable persons accept them; the same condemnation applies to the Kantian's single-minded emphasis on, and peculiar interpretation of, respect. Boettcher tries to blunt the force of this criticism by noting that many different comprehensive views insist upon treating others as ends in themselves. He's right, of course, but makes an irrelevant point. Boettcher needs to show not merely that many reasonable views incorporate Kantian respect to some degree, but that *all* reasonable views give *absolute* priority to a roughly Kantian interpretation of respect on *all* occasions.

Charles Larmore has offered a similar answer to this objection. According to Larmore, a standard of political legitimacy should be “freestanding” in the sense that it presupposes only a very “minimal morality, which reasonable people can share despite their expectably divergent religious and ethical convictions.”¹⁴ To his credit, Larmore goes on to show that the public reason liberal’s criterion of legitimacy depends upon a particular understanding of what it means to respect those subject to coercive laws. He thus embraces the conclusion that political liberalism rests upon the moral value of respect, where respect is interpreted in a roughly Kantian fashion. But doesn’t that mean political liberalism is not freestanding? Larmore answers,

I do think it important that the principle of respect does not express or entail a comprehensive moral philosophy. It has its place in a great many otherwise disparate ideas of the human good.¹⁵

Larmore is surely correct that public reason’s ideal of respect is a part of many different comprehensive views. But he makes the same mistake as Boettcher. It is one thing for a roughly Kantian respect to figure somehow into one’s comprehensive view; it is quite another to prioritize Kantian respect over all other political values on all occasions. Public reason liberals—who are often themselves sympathetic to Kantian and Rawlsian ideas—often fail to realize just how bold their criterion of legitimacy really is. Theirs is not the mundane claim that Kantian respect carries some importance for politics. They rather assert that their conception of respect is the political *summum bonum*. And that is obviously incompatible with the claim that political liberalism is freestanding.

There is, as I mentioned, a second strategy for defending the priority of justification-to. Rather than defending justification-to by an appeal to respect, the second answer denies the way I set up the problem in the first place. I claimed, recall, that reasonable people will affirm a diverse plurality of political values (including the value of justification-to). Further, I claimed that reasonable people will rank those values differently. The second strategy is to define “reasonable” in such a way that all the reasonable people agree that justification-to should be prioritized on every occasion. Many orthodox Rawlsians, for example, understand the term “reasonable” in a technical fashion.¹⁶

But I cannot, frankly, see how this second strategy can succeed either. Suppose I claim that all reasonable people regard Rawls as the greatest philosopher. Puzzled, you point out seemingly reasonable people who hold Plato, Aristotle, Karl Marx, Wilfred Sellars, or whomever in higher esteem. Undaunted, I explain that I don’t mean what you do by the word “reasonable.” I have invented my own technical definition, and according to it, my claim holds. You would in this case be quite right to complain that my assertion is either false or trivial. It is false if I use “reasonable” according to anything like the standards of natural language. After all, many of my fellow philosophers—smart people who sincerely consider their opponent’s views and change their minds when presented with contrary evidence—do not regard Rawls as the greatest. On the other hand, my assertion is trivial if I have concocted my own definition

of “reasonable” such that Rawls’s surpassing greatness follows analytically. And here’s the problem: whether my assertion is false or trivial, nothing much follows for your beliefs about Rawls’s greatness. Public reason liberals, it seems to me, pull nearly the same trick. They boldly claim that all reasonable people agree upon the priority of justification-to. But their claim is false if they mean to use “reasonable” in anything like its natural language sense. Their claim is trivial if they secure the agreement of the reasonable by definitional fiat. And whether their claim is false or trivial, nothing much follows about how the rest of us ought to regard the priority of justification-to.

If these criticisms are sound, public reason liberalism is, in its current form, deeply flawed. But I think it would be a mistake to abandon the public reason liberal’s project altogether. Justification-to is a *bona fide* political value—that, again, is the lesson of Estlund’s pope. And if we are going to be coerced, as we inevitably will be, I think we all would prefer it to be for reasons that we can see as legitimate. So the problem with public reason is *not* that it values justification-to. The problem resides instead in public reason’s simplistic account of the relationship between legitimacy and justification-to: a law is legitimate if and only if it satisfies the demands of justification-to. Justification-to is the political *summum bonum*. Period. Public reason liberals have developed sophisticated interpretations of what it is to justify a law in terms of reasons that all can accept; their reduction of legitimacy to justification-to is comparatively blunt. So rather than abandoning justification-to altogether, what we need is a more nuanced account of justification-to’s political significance.

I think that the legitimacy of any given legal rule is a function of several political values, including justification-to, and that the importance of justification-to varies from case to case. Sometimes, failing the demands of justification-to counts heavily against legitimacy, and other times it does not. I cannot in this short space develop a full theory of legitimacy. Instead, I will argue that justification-to’s importance to legitimacy varies (perhaps non-linearly) with two factors. First, justification-to becomes more important to legitimacy as the rule in question threatens more dire coercion. Second, the importance of justification-to also depends upon whether there is an adequate, less coercive alternative to the rule in question. I will spend the balance of this section defending these two claims.

Why does justification-to become more important to legitimacy as laws threaten more dire coercion? To answer, let us briefly reflect upon *why* public reason liberals emphasize justification-to in the first place. The public reason liberal’s ambition, as I understand it, is to justify law’s infringement upon liberty. Consent is perhaps the gold standard for reconciling liberty with legal coercion: by voluntarily agreeing to a coercive policy, we would waive whatever claims of liberty the policy would otherwise have violated. But consent theory fails because most citizens have not consented in the morally relevant sense of that term. Public reason liberals aim to forge a different connection to citizens’ wills. We cannot reach the gold standard of consent, but perhaps the silver standard is good enough: if laws can be justified to citizens in terms

of reasons that they themselves accept, we will have gone some distance towards ameliorating the coercion-liberty conflict. We might resent coercion without consent, but if we must be coerced, we will certainly prefer coercion that we can see as justified.

My point is that justification-to is a value, in part, *because* law is coercive. But if that is correct, the importance of justification-to should depend upon just how coercive the law under consideration happens to be. Compare: advanced weightlifters use protective gear to protect their joints during heavy lifts; that protective gear becomes more important as they lift heavier and heavier loads. Public reason liberals use justification-to in order to assuage the conflict between liberty and coercion; justification-to should therefore become more important as that conflict becomes more and more severe.

To see how plausible this suggestion is, consider a pair of examples. Suppose that a certain economic policy, which is only mildly coercive, would lay the foundation for universal, long-term economic prosperity. Unfortunately, however, we cannot justify it to all citizens—some are reasonably concerned that our public culture has already become too materialistic, that this new economic policy will worsen the trend, and that other political values therefore make an overwhelmingly stronger claim on our resources. An economic policy like this is perhaps unrealistic, but I am using a thought experiment to draw out an important point: while it would be *better* if we could somehow justify our policy to those who reasonably object, it is implausible that its failure to meet the demands of justifiability-to renders our policy illegitimate. There is admittedly some harm done to the objecting anti-materialists, harm that could be avoided by prioritizing justification-to. But since our economic policy imposes so little coercion, the harm to the small number of anti-materialists pales in comparison to the enormous value for everyone else (the anti-materialists can even devote their newfound prosperity to combatting the perceived ills of materialism). When the objecting minority would suffer so little, we should not let them hold everyone else's economic well-being hostage.

Now consider a contrasting case: a society is deciding whether to establish a prison system. Whereas the economic policy threatens only mild coercion, a prison system means that we, or our loved ones, can be locked away for years, or perhaps even for life. Worse, every single human-run system will make mistakes, so it is virtually certain that some citizens, or their loved ones, will be locked away without adequate cause. And whereas the economic policy will benefit everyone in objective terms, a prison system does not. One might initially think that a prison system will benefit everyone by providing the good of public safety, but the thought does not withstand scrutiny. Leave aside the substantial worry that prisons cause at least as much criminality as they prevent. Those who are incarcerated unjustly clearly do not benefit overall. Neither do their families. Others, those who live on isolated and peaceful communes, for example, do not much benefit either. Nor does anyone for whom the tax burden of supporting prisons exceeds the harm they would otherwise have experienced at the hands of criminals. So whereas the economic policy

realizes a great benefit for all citizens at the cost of very little coercion, the prison system threatens to direly coerce people who will not necessarily reap its benefits.

Now, I do not for a moment mean to suggest that the wrong of unjust imprisonment is somehow compensated for in advance if the prison system can be supported by reasons that all citizens accept. *That* would be an extraordinarily implausible claim. My point is that whereas it would not be wrong (or at least, not very wrong) to institute the economic policy, it *would* be wrong to institute the prison system without first satisfying the demands of justification-to. The prison system puts at least some citizens at risk of a dire liberty violation. Worse, it puts those citizens at risk for the sake of a benefit that they themselves will not partake of (it is, again, absurd to say that prisons benefit everyone). Given these facts, I think it clear that we have a duty to secure citizens' approval before instituting a prison system. Getting their explicit consent would probably be the best option, but making sure that the prison system is justifiable in terms of reasons that citizens would accept is, as I said, the silver standard. It is one step towards reconciling liberty with coercion, and given the level of coercion involved, it is a step we ought to take. Thus, it is *prima facie* plausible that the importance of justification-to varies with the degree of coercion a law threatens.

My second claim is that the importance of justification-to also depends upon whether there is an adequate, less coercive alternative to the rule in question. Consider, for a mundane example, the rule that campers may not cook over open fires during the dry season in particular regions of the country. The rule prevents campers from doing something they quite reasonably enjoy, and that restriction on liberty is a bad-making property of the rule. But the badness is more than compensated by the fact that the rule has proven *necessary* for preventing forest fires that devastate national parks, destroy citizens' property, and take fire fighters' lives. In this case, the necessity of the rule diminishes the importance of justification-to.

One might think that such a rule would satisfy the demands of justification-to rather easily—what reasonable person, after all, would object to a regulation that is necessary for such important things? The answer depends upon what, precisely, one means by reasonable. On a sufficiently thin conception of reasonableness, reasonable persons might object to the seasonal prohibition on campfires. On a thicker interpretation, they probably wouldn't. Defining "reasonable" is a treacherous business, and, fortunately, I don't need to pursue it. For my purposes, the following claim is sufficient: whether or not reasonable people will object, their objections count less heavily against legitimacy when there is no less coercive avenue to the relevant, overridingly important ends.

When, however, there is a less coercive avenue to the relevant ends, the necessity defense fails. If we could prevent forest fires equally well by installing specialized fire pits, citizens should have the right to say whether they would prefer to pay the modest price for the pits or to live with a seasonal ban on campfires. Given that there is a less coercive route to preventing forest fires

(suppose the cost of installing fire pits is so low that the extra tax burden amounts to negligible coercion), citizens' objections *do* count against the legitimacy of the campfire ban. Justification-to, in other words, is more important to legitimacy when there are less coercive alternatives to the rule in question.

To sum up: there is a plurality of political values, and I think that political legitimacy is a function of them all. But the function is complicated. The weight each value has in the calculation of overall legitimacy depends on the nature of the law in question. I have argued that the weight of justification-to is a function of two things: (a) the degree of coercion a rule threatens, and (b) the presence of adequate, less coercive alternatives.

5. RETRIBUTIVISM CANNOT MAKE LEGAL PUNISHMENT LEGITIMATE

I opened this paper with a dilemma from Hugo Bedeau. Bedeau charged that in the process of trying to explain why it is good that criminals suffer, retributivists would fall into one of two traps: they would either pound the table and insist upon retributivism's truth or else would cite consequentialist values in defense of punishment. But, as I mentioned in the introduction, there is too much space between this dilemma's horns: Bedeau does not explain why retributivists cannot cite deontological rather than consequentialist values.

The pieces I need to rehabilitate Bedeau's dilemma are now in place. We saw that legal punishment is but one way to respond to crime: whereas restorative justice aims to reinstate prior conditions, punishment aims to make criminals suffer to the "appropriate" degree. We saw that to count as a retributivist, one must answer the "why" question by saying that punishment is good because guilt calls for punishment, not for restoration. And we have seen that the importance of justification-to varies (a) with the degree of coercion a law threatens and (b) with the presence of adequate, less coercive alternatives.¹⁷

To begin putting these pieces together, notice that the laws establishing certain legal punishments (prisons, heavy fines, and execution) are precisely the kind of laws that need to satisfy the demands of justification-to. They (a) coerce direly and (b) there are adequate, less coercive alternatives. Against (b), one commonly hears the argument that legal punishment is necessary, but the necessity argument for legal punishment fails. After all, restorative justice procedures constitute an alternative to legal punishment, and evidence suggests it is an alternative that safeguards society at least as well as punishment does.¹⁸ We need a mechanism for dealing with crime, true enough. But the necessity argument illicitly assumes that punishment is the only viable mechanism. Indeed, there is an impressive literature on alternatives to punishment.¹⁹

So the laws establishing legal punishment need to be justified in terms of reasons that citizens can accept. But what, exactly, does it mean to justify a rule in terms of reasons that another person accepts? There are two ways to parse this idea, a stronger and a weaker interpretation of

justification-to. On the stronger interpretation, justification-to means showing that, of all the reasons the other accepts, the set that favors the rule outweighs the set that disfavors the rule. Or, more pithily, strong justification-to consists in showing that adopting the rule is an all-things-considered ought relative to what the other regards as reasons.

Whereas strong justification-to takes the perspective of an all-things-considered ought, weak justification-to shows the other merely that she endorses at least one reason that favors the rule. Weak justification-to is thus very weak. It is probably too weak to assuage the conflict between liberty and coercion—that a single reason of yours favors a given law is small comfort when the overwhelming balance of your reasons condemns it—and thus too low a standard for political legitimacy. Weak justification-to is nonetheless useful for my purposes. These days, most philosophers who support legal punishment do so on retributivist grounds. But if they hope to make legal punishment legitimate, their retributivist arguments must satisfy the demands of justification-to. I will show that retributivist arguments do not even satisfy the demands of weak justification-to.

The problem is that there are, among the reasonable, people who do not see guilt as calling for punishment. I do not see guilt that way. Nor does David Boonin, Geoffrey Sayre-McCord, Desmond Tutu, or Nelson Mandela.²⁰ Certain religious traditions do not see guilt as calling for punishment. And to cut short what would otherwise be a very long list, nor does *any* advocate of restorative justice. Retributivists therefore need to *show* that guilt calls for punishment, and they must show that in terms that non-retributivists accept. It is at this point that a variation of Bedeau's dilemma applies. Any explanation of why guilt demands punishment will fall into one of two traps: it will either take the truth of retributivism as a premise and thus fail as justification-to (indeed, even as weak justification-to), or else it will invoke non-retributive values, in which case retributivism ceases to function as a justification of punishment. In the balance of this section, I will consider the horns of this dilemma one at a time.

A. THE FIRST HORN: CIRCULARITY

A surprising number of retributivists attempt to show that guilt demands punishment by appealing to the argument from cases.²¹ The argument begins by presenting cases in which, the retributivist hopes, we will find irresistible the conclusion that the relevant criminals ought to be punished (the "relevant criminals" are typically guilty of particularly heinous crimes). The argument treats retributive judgments about such cases as data points and concludes that retributivism is the principle that best fits them. This argument is, perhaps, effective rhetoric. But it fails badly *qua* justification-to. In fact, it fails badly *qua* weak justification-to. If our judgments about particular cases were independent of our general conviction that guilt calls out for punishment, the argument would work. The problem, however, is that it is extremely implausible that our general retributive intuitions do not color our judgment in particular cases. Our general retributive intuitions contribute to particular retributive judgments, which the argument from cases cites, in turn, as evidence

of retributivism. The argument therefore gives those who distrust their general retributive intuitions (or lack them altogether) no reason at all to endorse retributivism.

Michael Moore presents a subtler defense of retributivism by appealing to virtuous emotion. His initial premise is that a virtuous person's emotions are a reliable, albeit defeasible, guide to truth—a virtuous person's disgust at racism, for example, is good evidence that racism is immoral. Let us grant Moore this premise for the sake of argument. He then asks the reader (who he presumes to be virtuous) to imagine how she would feel if she had committed an awful murder. Moore thinks that even if we undertook whatever restitution is possible, we would not object to the suffering entailed by our punishment. "Our feelings of guilt," he writes, "thus generate a judgment that we deserve the suffering that is punishment."²² The last step in his argument is that we should not hesitate to generalize from our own case to the cases of others. It would be "elitist and condescending towards others not to grant them the same responsibility and desert you grant to yourself."²³ If our guilt calls for punishment, so, Moore concludes, does the guilt of others.

But Moore's argument, like the argument from cases, is circular, albeit less obviously so. The problem is that he simply *assumes* that it is inconsistent with virtue to experience guilt as calling for restoration rather than punishment. But that is, of course, to assume the point that needs proving. Those who doubt that guilt calls for punishment in the first place will be entirely unmoved.

Shaun Nichols has recently tried to grab hold of the circularity horn by arguing that even *if* justifications of retributivism are circular, we have an excellent reason not to jettison the notion that guilt calls for punishment.²⁴ His argument is sophisticated, and if sound, would satisfy the demands of justification-to in an unexpected way: by showing that non-retributivists are mistaken to ask for a justification in the first place.

Somewhat surprisingly, Nichols's argument begins by admitting that we can tell the sort of genealogical story about retributivism that seems, *prima facie*, capable of debunking it. Nichols's genealogy is best motivated by a puzzle: Cultures develop many different norms. Some fall out of favor; others stick with us. What explains the persistence of some norms and the disappearance of others?

According to Nichols, most human beings are disposed to similar affective responses. Affective response is, of course, not entirely uniform across people, but there is great overlap—many of the things that prompt sympathy, jealousy, disgust, or anger in us are likely to elicit similar responses from others. Some cultural norms "resonate" with these natural affective responses.²⁵ We feel anger, for example, when someone harms us, or our loved ones, by transgressing accepted rules. Anger makes us want to harm those who harm us, so retributivism holds obvious appeal. Indeed, Nichols presents a series of experiments that strongly suggest that we turn to retributive norms when we become angry at a perceived infraction.²⁶ Now, retributivists usually emphasize that retribution is not the same thing as

vengeance—whereas retribution is a measured response to crime, vengeance is unruly and often goes overboard. But Nichols does not deny the distinction between retribution and vengeance; he makes the plausible suggestion that retributive norms have proven durable, in part, because they tap into emotions that come naturally to us.

Prima facie, Nichols's genealogical account debunks retributivism. If Freud is correct that belief in God stems from wishful thinking, we should have little confidence that God exists—wishful thinking is, after all, epistemically unreliable. If Nichols is correct that retributivism is a cultural artifact sustained by anger, we should have little confidence that retributivism is true—anger is arguably a-rational and thus epistemically unreliable.

One might object that Nichols's genealogical account needs some "thickening." After all, retributive norms are sustained by a host of factors in addition to anger. For example, retributive ideas are embedded in traditional western theology as a part of the crucifixion story—Jesus's death was necessary to satiate the demands of our guilt. Retributivism is thus sustained, in part, by people's commitment to a particular religious way of life. And religion does not stand alone: retributivism is also embedded in the law, in political parties, and in other social institutions. But the addition of social institutions to Nichols's genealogy does little to strengthen retributivism. As Socrates would ask Euthyphro: Is retributivism justified because it's embedded in certain social institutions, or do social institutions tend towards retributive ideas because the ideas themselves are justified? The first option courts an implausible relativism; the second robs institutional embedded-ness of any power to justify retributivism. So, again, it looks as if both Nichols's genealogy and its thicker cousin threaten to debunk retributivism. But Nichols resists that conclusion by endorsing a view he calls ethical conservatism.

Ethical conservatism claims that if a norm is (i) widespread, (ii) inferentially basic, and (iii) resonates with common human emotions, it needs no justification to carry normative weight.²⁷ The properties of being widespread and resonating with common human emotions require no further explication. But what makes a norm inferentially basic? The answer is that inferentially basic norms "are not the result of consciously available inferences from other norms or facts."²⁸ On reflection, we see the norm that one ought not fire guns at innocent persons as a consequence of the more general norm that prohibits murder. The conception of guilt-that-calls-for-punishment, by contrast, does not seem reducible to anything else. As Nichols aptly puts it, retributivism is "an independent part of our moral worldview."²⁹

With this background in place, Nichols's argument is straightforward.

- (1) Ethical conservatism is true.
- (2) Retributivism is a norm that is (i) widespread, (ii) inferentially basic, and (iii) resonates with natural human emotions.

Therefore,

(3) Retributivism does not need a justification to carry normative weight.

The argument is valid, and premise (1) does most of the heavy lifting. So why think that ethical conservatism is true? Nichols identifies two ways to justify a norm: (a) show that it belongs to a class of ultimately justified norms, or (b) show that it is derived from ultimately justified norms. His argument is that neither path to justification will work for inferentially basic norms, and that without inferentially basic norms, we would end up with an untenable “emaciated ethics.”³⁰ All inferentially basic norms—being inferentially basic—are not derived, so option (b) is out. And Nichols is, rightly, I think, skeptical of option (a). There is ample room to doubt the Kantian’s derivation of norms from reason, the intuitionist’s claim to ascertain the truth of moral norms a priori, and so forth. Since neither (a) nor (b) is a likely path to justification for inferentially basic norms, rejecting ethical conservatism would entail an emaciated ethics.

Nichols’s argument, here, is too quick. His defense of premise (1) presumes a problematic sort of foundationalism: norms are justified if they (a) belong to a class of ultimately justified norms, or (b) are consciously inferred from ultimately justified norms. The problem is that there are other paths to justification. Here’s one: a belief is *prima facie* justified if it coheres with other apparently reliable beliefs. One needn’t be a full-blown coherentist about justification to agree with this principle. This principle is available to, at least, coherentists, reliabilists, and pragmatists. (I think just about every plausible epistemology will say that coherence with apparently reliable beliefs constitutes *some* evidence of truth.) Therefore, Nichols is not entitled to premise (1), and has given us no reason to think that retributivism does not require justification.

But Nichols *has* arguably given us a reason to doubt retributivism. He began, remember, with the plausible suggestion that our retributive norms and intuitions are sustained, in part, by anger. So far from showing that retributivism stands on sure footing, Nichols has put the retributivist in an even worse position: she must not only convince her critics that guilt calls for punishment, she, arguably, must explain how anger, which is a-rational, somehow managed to preserve true intuitions.

If Nichols’s argument had worked, he would have satisfied the demands of justification-to by showing that, contrary to initial appearances, it is inappropriate to demand an explanation of why guilt calls for punishment. But his argument fails, so the requirement of justification-to stands. Circular arguments, like Moore’s or the argument from cases, fail to satisfy that requirement. In fact, they fail to satisfy the much weaker requirement of weak justification-to.

B. THE SECOND HORN: CITING NON-RETRIBUTIVIST VALUES

Bedeau rightly pointed out that a defense of retributivism would fail if it cited consequentialist values: a

consequentialist argument might justify punishment, but if it does, the justification would not be retributivist in character. Notice, though, that citing consequentialist values is only the most obvious way to impale oneself on the second horn of my dilemma. The other is to cite deontological values that have nothing to do with retributivism.

Consider, for example, defenses of retributivism that emphasize the moral concept of *fairness*. The idea behind such views is that by breaking legal rules, criminals unfairly take advantage of their fellow citizens’ legal obedience. John Finnis, for example, says that criminals “indulge in a (wrongful) self-preference.”³¹ Punishment is important, on these views, to restore fairness—punishment offsets the extra liberty or discretion criminals take for themselves.

There is much to *dislike* about fairness-based views. Most obviously, it is often implausible that a breach of fairness is the wrong-maker. As many have observed, rape, sexual abuse, assault, battery, manslaughter, murder, theft, and the like are principally wrong because they harm the victim. So if we punish rapists, abusers, or murderers because it is unfair of them to break the rules that the rest of us obey, we would punish them for the wrong reason. My criticism is that, substantive flaws aside, invoking the deontological value of fairness would make the retributivist’s concept of guilt-that-calls-for-punishment otiose.

The first job for a theory of punishment, recall, is to answer the “why” question—to explain why it is morally important that criminals suffer punishment. It is a common, but unfortunate, practice to lump all backward looking answers together under the retributivist banner.³² A more perspicuous classification would sort defenses of punishment by the moral concept they invoke to answer “why.” What I have been calling retributivism invokes the concept of guilt-that-calls-for-punishment. A fairness-based theory, by contrast, answers the “why” question by invoking the concept of fairness: a criminal should be punished to offset whatever “extra” privileges she unfairly claimed as her own. The retributivist’s peculiar conception of guilt shows up nowhere in that answer. So by citing fairness, the retributivist would simply abandon retributivism.

It is important to see the problem, here, in terms of weak justification-to. The retributivist’s task is to show those of us who do not share her view that guilt does, in fact, demand punishment. But as Bedeau saw, she cannot convince reasonable skeptics of retributivism by citing consequentialist values to answer the “why” question; my point is that she cannot convince reasonable skeptics by citing non-retributive deontological values like fairness, either. So even if a fairness-based theory can justify punishment (I am skeptical about that too), fairness gives non-retributivists *no reason at all* to endorse punishment on retributivist grounds.

One might object that while retributivism and fairness are different answers to “why,” the answers are complimentary. We can form a retributive-fairness hybrid: criminals should be punished because they are guilty; but we can use fairness to define guilt. This sort of hybrid unfolds in two

steps. First, we use the concept of fairness to make sense of the concept of guilt—the guilty are precisely those who breach fair rules, and they deserve punishment in proportion to their breach of fairness. Second, and along with the retributivist, we insist that guilt calls for punishment.

The problem, however, is that retributivism is still, from the perspective of justification-to, an idle third wheel in this retributive-fairness hybrid. In order to count as something new, the hybrid would have to answer the “why” question differently than a pure fairness theory. But it doesn’t: the hybrid holds that criminals should be punished *because* their behavior is an offense against fairness, and that they should be punished *in proportion* to the degree that their behavior is unfair. Thus, the retributivist’s peculiar concept of guilt-that-calls-for-punishment has been expunged, without loss, from the hybrid’s answer to “why.” Again, it is important to see the problem from the perspective of justification-to. The retributivist cannot defend her own answer to “why” by citing an altogether different answer.

It should, at this point, be clear that this is a general problem. If a retributivist’s explanation of why it is good to punish incorporates any moral concept other than guilt-that-demands punishment, she is open to the criticism that criminal desert no longer plays any necessary role in her view. Of course, we have already seen what happens to views that do not incorporate other moral concepts: their circularity means they fail as justification-to, even as weak justification-to. Bedeau’s dilemma is therefore rehabilitated. Any explanation of why guilt calls for punishment will (a) presuppose the very retributive intuitions that are at issue, or (b) cite non-retributive values. Either way, retributivism fails as justification-to. Given that the laws establishing legal punishments like prisons, execution, or heavy fines are precisely the kind of law that need to satisfy the demands of justification-to, we get the conclusion that retributivism cannot make legal punishment legitimate.

6. CONCLUSION

I would like to close by making one clarification and answering one objection. The clarification is that, as I stated in the introduction, my aim has been to show that retributivism cannot make harsh punishments like imprisonment or execution legitimate. As a matter of personal belief, I do not think that retributivism can make lesser punishments legitimate either—I think retributivism is false and therefore incapable of justifying anything. But my argument in this paper does not depend upon the assumption that retributivism is false. It notes, instead, that justification-to is especially important to the legitimacy of imprisonment, heavy fines, and execution, and that Bedeau’s dilemma, when properly rehabilitated, shows that retributivism fails as justification-to.

The objection I have in mind claims that my argument proves too much. My recasting of public reason implies that every defense of legal punishment is accountable to the burdens of justification-to. Retributivism fails, but so, the objection goes, will every other defense of punishment. Thus, the objection concludes, my argument threatens to condemn the whole institution of legal punishment.

My response to this objection is two-fold. First, I am not sure that every defense of legal punishment will fail. Retributivism fails as justification-to because its defense of punishment is so flat-footed: it’s good to punish criminals because they deserve it. There is no a priori reason why other defenses of punishment cannot be better. Second, and more importantly, I do not think we should shy away from the conclusion that legal punishment is an illegitimate practice. To assume that punishment must be justifiable would be to prejudice inquiry at the outset.

But perhaps the force of this objection is rather that my recasting of public reason threatens to condemn not only legal punishment, but also every method for responding to crime. After all, for any arrangement that aims to solve the social problems posed by crime, there is likely to be at least one reasonable objector. *This* objection, however, is too hasty. Justification-to becomes more and more important to legitimacy as the rule in question threatens more and more coercion. Compared to legal punishment, a practice built around reparation or restoration is much less coercive. Such a practice would not necessarily have to meet the burdens of justification-to; the substantial political values such a program would serve may be sufficient.³³

ACKNOWLEDGEMENTS

I would like to thank Nate Olson, Katherine Schweitzer, Nate Budolfson, Jake Earl, Sam Boldenow, Bryce Huebner, George Rudebusch, and Marcus Hedahl for comments on earlier drafts of this paper.

NOTES

1. For a nice summary, and example, of this literature, see David Boonin, *The Problem of Punishment* (New York: Cambridge University Press 2008), 25.
2. H. L. A. Hart, “Prolegomenon for the Principles of Punishment,” Presidential Address to the Aristotelian Society, 1959, reprinted in his *Punishment and Responsibility* (Oxford: Clarendon Press, 1968).
3. Boonin, *The Problem of Punishment*, 24-25.
4. Richard Wasserstrom, “Capital Punishment as Punishment: Some Theoretical Issues and Objections,” *Midwest Studies in Philosophy* 7 (1982): 476; Thom Brooks, *Punishment* (New York: Routledge, 2012), 1-2.
5. Anthony Duff, “Restorative Punishment and Punitive Restoration,” in *A Restorative Justice Reader* (Portland, OR: Willian Publishing 2003), 382.
6. *Ibid.*, 393; emphasis in original.
7. One might object that on rehabilitative theories of punishment, punishment aims to educate—it is supposed to show criminals the wrong they do. So does my account of punishment imply that rehabilitative theories are false? My answer is a qualified “yes.” I answer “yes” because, so far as I can see, rehabilitation is just not punishment. The meth addict who goes to rehab is not punished, even if her addiction is own fault and even though rehab is painful. Knee-replacement surgery involves painful rehabilitation, but it makes no conceptual sense to say that the patient is punished, even if the injury is her own fault. My “yes” is qualified because I think it is an open question whether rehabilitation, rather than punishment, is the correct response to crime. If the rehabilitative theorist is offering an analysis of punishment, her view is implausible; if she wants to replace punishment with rehab, she has an interesting proposal.
8. Rawls famously split this question into two. Why is it important to have a practice of punishment? And, why is it important to punish on particular token occasions? Rawls used utilitarian considerations to answer the first question and more deontological ideas to answer the second. Like most, I think Rawls’s divide and conquer strategy exhibits the same

instabilities as rule utilitarianism. See John Rawls, "Two Concepts of Rules," *Philosophical Review* 64 (1995).

9. The search for an alternative to lex has been difficult, but the theoretical point remains. Nothing about the retributivist's answers to "why" and "who" commits her to citing lex in answer to "how."
10. David Enoch, "Against Public Reason" (forthcoming).
11. I am neutral, here, between a convergence account of public reason and a consensus account. According to convergence, each citizen can have her own unique reasons for accepting *L*. Consensus, by contrast, would require that each citizen accept *L* for the same reason.
12. David Estlund, *Democratic Authority* (Princeton: Princeton University Press, 2008), 5.
13. James W. Boettcher, "The Moral Status of Public Reason," *The Journal of Political Philosophy* 20 (2012): 176.
14. Charles Larmore, "The Moral Basis of Political Liberalism," *The Journal of Philosophy* 96 (1999): 600.
15. *Ibid.*, 623.
16. For a nice discussion of how Rawls uses "reasonable," see Samuel Freeman, *Rawls* (New York: Routledge, 2007), 295–300.
17. If public reason liberals are unconvinced by my arguments in the last section, they are welcome to regard justification-to as the political *summum bonum*. That assumption is more than sufficient for my argument in this section. But I hope to have convinced those who share my suspicion of public reason that we should not abandon justification-to as a political value.
18. Brooks summarizes and cites the evidence in Brooks, *Punishment*, 83–84.
19. See Boonin, *The Problem of Punishment*, 213–69, for a summary.
20. *Ibid.* Geoffrey Sayre-McCord, "Criminal Justice and Legal Reparations as an Alternative to Punishment," *Philosophical Issues* (2001): 502–29.
21. Boonin, *The Problem of Punishment*, 87–93.
22. Michael S. Moore, "The Moral Worth of Retribution," in *Responsibility, Character, and the Emotions*, ed. Ferdinand Schoeman (New York: Cambridge University Press, 1987), 215.
23. *Ibid.*
24. Shaun Nichols, "Brute Retributivism," in *The Future of Punishment*, ed. T. Nadelhoffer (New York: Oxford University Press, 2011).
25. *Ibid.*, 26–32.
26. *Ibid.*, 27–32.
27. *Ibid.*, 38.
28. *Ibid.*, 31.
29. *Ibid.*, 32.
30. *Ibid.*, 38.
31. John Finnis, "The Restoration of Retribution," *Analysis* 32 (1972): 132.
32. Boonin (*The Problem of Punishment*) stands as a partial counter example to this trend. He calls all backward-looking views retributivist, but distinguishes fairness-retributivism from desert-retributivism. I have argued that desert-retributivism is the only view that deserves the retributive title.
33. See note 17.