Retributivists! The harm principle is not for you!

Article

Published Version

Retributivists! The Harm Principle Is Not for You!*

Patrick Tomlin

Retributivism is often explicitly or implicitly assumed to be compatible with the harm principle, since the harm principle (in some guises) concerns the content of the criminal law, while retributivism concerns the punishment of those that break the law. In this essay I show that retributivism should not be endorsed alongside any version of the harm principle. In fact, retributivists should reject all attempts to see the criminal law only through (other) person-affecting concepts or "grievance" morality, since they should endorse the criminalization of conduct that is either purely self-harming or good for somebody and bad for nobody (i.e., Pareto improvements).

INTRODUCTION

There are, to (over)simplify, two main questions within the philosophy of criminal law: What (if anything) justifies state punishment? And what kinds of conduct (if any) are (or may be made) punishable—what can or should we criminalize? Retributivism is a theory of punishment. In justifying punishment, it focuses our attention not on the future benefits that punishment may bring but rather on past wrongdoing, in virtue of which, so it is claimed, punishment is deserved. The harm principle is a

* I began thinking about retributivism and its relationship to harm and harm prevention when I worked on the AHRC-funded Preventive Justice project (AH/H015655/1) with Andrew Ashworth and Lucia Zedner. I am grateful to Andrew and Lucia, and to Antony Duff, Keith Hyams, Catriona McKinnon, Victor Tadros, and the editors and referees of *Ethics* for written comments. The article has benefited from discussion at the University of Warwick and the University of Reading. I should perhaps note, given the subject matter of the article, that I am not a retributivist, or at least not a committed one. But I am interested in the structure and implications of the view, especially for other areas of criminal law theory, and the present article is written in that spirit.

Ethics 124 (January 2014): 272–298

popular principle of criminalization.[1] As I will show, there are a variety of ways to understand the harm principle, but each demands that when we enact a criminal law we focus on harmful conduct, conduct that risks harm, and/or the prevention of harm. Since retributivism has been revived in recent years, and since debates concerning the moral limits of the criminal law continue to be conducted through the lens of the Hart-Devlin debate, and Feinberg's account, retributivism and the harm principle are widely accepted positions, and perhaps the dominant positions, within punitive and criminalization theory, respectively.[2]

In this article, I show that these two commonly held positions do not make good bedfellows: retributivists should not accept any version of the harm principle. For some versions of the harm principle, this is because retributivism and the harm principle (or its grounds) are simply logically incompatible. For other versions of the harm principle, this is because, while the positions are logically compatible, the retributivist can only endorse the harm principle at the expense of adopting highly implausible positions concerning the content of the criminal law, denying the criminal law the ability to protect the criminal justice process itself—a key element of the rule of law. Furthermore, the positions the retributivist would be required to adopt in order to endorse the harm principle contradict the importance she places on doing retributive justice and the central place that retributive justice occupies within her understanding of the importance and justification of the criminal law. My aim is to show that retributivism and all versions of the harm principle are unattractive in combination.

RETRIBUTIVISM AND THE HARM PRINCIPLE

If I am able show that retributivists should not accept any version of the harm principle, then this is an important finding: two of the main positions within criminal law theory should not be endorsed alongside one another. By and large, criminalization and punishment theories are discussed in isolation from one another, leading to a latent assumption that we can pick and choose from among theories in either debate without that leading to consequences for our stance in the other. At first blush,

1. Or, at least, it is often thought of as such. I will show below that Mill's harm principle extends far beyond the content of the criminal law.

2. On the rise of retributivism, see David Dolinko, "Three Mistakes of Retributivism," *UCLA Law Review* 39 (1991–92): 1623–57 at 1623–24; Michael Davis, "Punishment Theory's Golden Half Century: A Survey of Developments from (about) 1957 to 2007," *Journal of Ethics* 13 (2009): 73–100 (although Davis disputes that all the theories that have been called retributive are best thought of as such). As the article proceeds, the number of influential figures in the philosophy of criminal law who have endorsed versions of the harm principle, often alongside retributivism, will become apparent.

this seems correct—retributivism and the harm principle may seem fully independent, as they are theories, or parts of theories, about different things: punishment and criminalization. This point is emphasised by Jean Hampton, who states that "there is no inconsistency between the harm principle and the principle of retribution . . . because the harm principle is a theory of the content of legitimate law, and the retributive principle is a theory of punishment for (legitimate) law breakers."[3]

Where these theories have been discussed alongside one another, several theorists have assumed their compatibility or have even argued that they are mutually supporting. Hugo Bedau talks of "two basic retributive principles: (1) the severity of the punishment must be proportional to the gravity of the offence, and (2) the gravity of the offence must be a function of fault in the offender and the harm caused to the victim,"[4] suggesting that only harmful conduct is to be punished. Similarly, Robert Nozick claims that deserved punishment is a function of harm (done or intended) and the responsibility of the offender.[5] Hampton believes that the two positions support one another, arguing that both "rights-based" liberalism and retributivism have a similar basis. "Indeed," she writes, "it is surely no accident that the clearest proponent of retribution in philosophical history, namely Kant, is also one of the most ardent supporters of the liberal conception of the state." According to Hampton, Kant's view is one in which retributive punishment "is properly responsive to [violations of] those individual rights that are also part of the moral foundation for the harm principle. Such a view makes the philosophical foundations of the harm principle and retribution not only consistent, but also mutually supportive."[6]

Yet, as I shall show, they are neither compatible in a package that is attractive, nor are they mutually supporting. This is not simply a matter of theoretical importance: some countries—for example, Finland and Sweden—have a legal culture which explicitly embraces both the harm

3. Jean Hampton, "How You Can Be Both a Liberal and a Retributivist," *Arizona Law Review* 37 (1995): 105–16, 106. See also Herbert Morris, "Professor Murphy on Liberalism and Retributivism," *Arizona Law Review* 37 (1995): 95–104.

4. Hugo Adam Bedau, "Classification-Based Sentencing: Some Conceptual and Ethical Problems," *New England Journal of Criminal and Civil Confinement* 10 (1984): 1–26, 13. Michael Davis uses this claim as the starting point for his rejection of harm as a central concept for retributivism in his "Harm and Retribution," *Philosophy and Public Affairs* 15 (1986): 236–66.

5. Robert Nozick, *Anarchy, State and Utopia* (New York: Basic Books, 1974), 60–62. Nozick was later happy for H in his $r \times H$ formula (where r is responsibility) to stand for "the wrongness or harm, done or intended, of the act" (*Philosophical Explanations* [Cambridge, MA: Harvard University Press, 1981], 363). The later Nozick leans toward the "wrongness" formulation (388–89) but does not reject the "harm" formulation and remains committed to the idea that all punishment-worthy action must have a victim.

6. Hampton, "How You Can Be," 115.

principle and retributivism (in that all criminal punishments must be in proportion to the harmfulness of the conduct).[7]

Some retributivists have resisted a harm-focused theory of criminalization. Yet nobody, so far as I know, has argued directly from the retributivist theory of punishment to the rejection of the harm principle in the way that I do here. The typical strategy is to show the retributivist some kinds of conduct which she will, it is hoped, believe warrant punishment but which cannot be criminalized according to the harm principle.[8] My argument, in contrast, is based on the harm principle's incompatibility with the value of retributive justice itself.[9] My claims are either claims of logical incompatibility or involve pointing to substantive criminal laws which we endorse, and which the retributivist must defend on the grounds of protecting the value of retributive justice, but which the harm principle would disallow. Therefore, my arguments are less easily brushed off by harm-focused retributivists. While such a retributivist may simply deny that dwarf tossing,[10] the desecration of graves, or the destruction of species[11] represent the kinds of wrongdoing that should attract the interest of the criminal law, she cannot deny the value of retributive justice, nor can she deny that that value is properly a concern of the criminal law. Ex hypothesi she has affirmed that value and has accepted that our states should use the criminal law and the criminal justice system to pursue that value.

In showing a deep incompatibility between retributivism and the harm principle, my argument represents a narrower but deeper contribution to the assessment of the harm principle than some recent critiques of it. It is narrower because it shows an incompatibility with a specific group of theories of punishment, which, while popular, are controver-

7. Essentially, the Swedish and Finnish sentencing regulations are legal instantiations of Nozick's formula (see n. 5 above). Andrew von Hirsch, *Censure and Sanctions* (Oxford: Clarendon, 1993), 90–91. I am grateful to Magnus Ulvang for discussion here.

8. See, e.g., Davis, "Harm and Retribution"; R. A. Duff, "Towards a Modest Legal Moralism," *Criminal Law and Philosophy* (forthcoming); Michael S. Moore, *Placing Blame* (Oxford: Clarendon, 1997), chap. 16.

9. Jeffrie Murphy offers a different argument for the incompatibility of retributive justice and the harm principle, claiming that the retributivist's commitment to punishing more for certain mental states experienced during harmful conduct reveals her to not only be concerned with harm, thus revealing a deep incompatibility between retributivism and the harm principle. I don't have the space to discuss Murphy's views in detail here, but I find them unpersuasive for reasons articulated, in particular, by Herbert Morris. See Jeffrie G. Murphy, "Legal Moralism and Liberalism," *Arizona Law Review* 37 (1995): 37–93, and "Legal Moralism and Retribution Revisited," *Criminal Law and Philosophy* 1 (2007): 5–20; Morris, "Professor Murphy on Liberalism and Retributivism."

10. Duff, "Towards a Modest Legal Moralism."

11. Moore, *Placing Blame*, 646.

sial, and therefore I am only able to show that the harm principle should be rejected if one is a retributivist. It is left open whether we ought to reject retributivism or the harm principle—but one or the other must be rejected. The argument is deeper, however, because if we wish to be, or remain, retributivists, then the argument presented will lead us much further afield from the harm principle than those who reject the harm principle often point us. I do not try to show merely that a retributive criminal law should expand its scope to include conduct that negatively affects other people in nonharmful ways, and therefore the argument does not point the staunch retributivist toward an additional or alternative person-affecting concept with which to limit criminalization—something (like offense or sovereignty) that is like harm but not harm.[12] Rather, retributivists, I claim, should reject all views of the criminal law that try to see it purely through the lens of "(other) person-affecting (political) morality."

RETRIBUTIVISM

Since I want to argue that the harm principle is incompatible with the value of retributive justice, I should say something about what this value is and how I understand it. Retributivism is a theory of punishment: it seeks to justify the normally impermissible treatment of individuals that comprise state punishment. There are many different types of retributivism,[13] but in justifying punishment all give a starring role to the concept of desert. According to R. A. Duff there is a "core retributivist thought: that what gives criminal punishment its meaning and the core of its normative justification is its relationship, not to any contingent future benefits that it might bring, but to the past crime for which it is imposed."[14] Elsewhere Duff writes that retributivism "justifies punishment in terms not of its contingently beneficial effects but of its *intrinsic* justice as a response to crime; the justificatory relationship holds between present punishment and past crime, not between present punishment and future effects."[15] To begin our discussion, any theory can be considered retributivist if it endorses this principle:

12. For critiques which take this tack, see Joel Feinberg, *Offense to Others* (New York: Oxford University Press, 1985); Arthur Ripstein, "Beyond the Harm Principle," *Philosophy and Public Affairs* 34 (2006): 215–45; Victor Tadros, "Harm, Sovereignty, and Prohibition," *Legal Theory* 17 (2011): 35–65.

13. John Cottingham, "Varieties of Retribution," *Philosophical Quarterly* 29 (1979): 238–46; Nigel Walker, "Even More Varieties of Retribution," *Philosophy* 74 (1999): 595–605.

14. R. A. Duff, "Retrieving Retributivism," in *Retributivism: Essays on Theory and Policy*, ed. Mark D. White (Oxford: Oxford University Press, 2011), 3–24, at 3.

15. R. A. Duff, *Punishment, Communication and Community* (Oxford: Oxford University Press, 2000), 19–20.

> The Retributive Principle: the state is (at least sometimes) justified in imposing punishment on past wrongdoers, even when punishing them will not prevent harm.

This may be both under- and overinclusive in terms of philosophers' self-descriptions as "retributivist."[16] For now, let us focus on the kind of "pure" retributivism as described by Duff (where preventive concerns play no role in the justification of punishment) and articulated by the Retributive Principle (which is a little less restrictive, as harm prevention is simply not a necessary condition of punishment, such that justified punishment which does not prevent harm is a conceptual possibility).[17] I will later show how anybody who endorses comparative proportionality in punishment should reject the harm principle, bringing more (perhaps all) self-identified retributivists within the scope of the article.

In deciding whether to punish the most serious crimes (e.g., murder), retributivists as described by the Retributive Principle will not need to know whether punishing the offender will prevent or deter future crimes—the importance of retributive justice (giving the offender the punishment he deserves) is, at least sometimes, enough to make punishment all-things-considered warranted, even in the absence of harm prevention. This will include the simple "because they deserve it" retributivism exemplified by Michael Moore[18] and retributivists who appeal to "fair play" considerations in justifying punishment.[19]

The Retributive Principle straddles both what I will call "telic" and "deontic" retributivism.[20] The telic retributivist's foundational belief is that when we punish people because they deserve it, we make things (in one way) better—retribution is good. The deontic retributivist's founda-

16. For example, it could include moral education theories, communication theories, and threat-based theories, provided that moral education, communication, or the realization of the threat are (at least sometimes) recommended even when doing so will not prevent harm. It excludes "retributivists" who see harm prevention as a necessary condition of punishment.

17. It is worth noting that the Retributive Principle differs from (and is more demanding than) the Retributivist Belief, around which I base my "Time and Retribution," *Law and Philosophy* (forthcoming), http://link.springer.com/article/10.1007%2Fs10982-013-9196-z.

18. Moore writes that "punishing the guilty achieves something good—namely, justice—and . . . reference to any other good consequences is simply beside the point" (*Placing Blame*, 111).

19. See, e.g., John Finnis, "The Restoration of Retribution," *Analysis* 32 (1972): 131–35; Herbert Morris, "Persons and Punishment," *Monist* 52 (1968): 475–501; Jeffrie Murphy, "Marxism and Retribution," *Philosophy and Public Affairs* 2 (1972): 217–43; Richard Dagger, "Playing Fair with Punishment," *Ethics* 103 (1993): 473–88; Davis, "Harm and Retribution."

20. Mitchell N. Berman, "Two Kinds of Retributivism," in *Philosophical Foundations of the Criminal Law*, ed. R. A. Duff and Stuart P. Green (New York: Oxford University Press, 2011).

tional claim is that we ought to punish those who deserve it, and this claim need not rest on a claim about value.

The Retributive Principle also captures "moral" and "legal" retributivists alike.[21] Moral retributivists believe that the wrongdoing described in the Retributive Principle is moral wrongdoing and that at least some moral wrongdoing is pre-institutionally defined. Legal retributivists believe that the wrongdoing in question is doing something that is criminally prohibited. Some legal retributivists may straightforwardly reject the harm principle. Since they believe that legally defined wrongdoing warrants punishment, if they also believe that the legislature has a free hand over what it makes criminal (and thus justly punishable), then the legal retributivist has already rejected the harm principle (and indeed any substantive principle) as a principle of criminalization. However, the legal retributivist may believe that while just punishment is connected to de facto criminalization, there are nevertheless moral limits on de jure criminalization. This kind of legal retributivist may be tempted to accept the harm principle as one such limit. I will show why she would be wrong to do so.

HARM PRINCIPLES

The harm principle is often presented as a theory (or, at least, one necessary condition) of criminalization: it explains what the criminal law should proscribe (or, at least, what it should not). It is a position closely identified with liberalism. According to A. P. Simester and Andreas von Hirsch, it is "a familiar tenet of liberalism that the state is justified in intervening coercively to regulate conduct only when that conduct causes or risks *harm* to *others*,"[22] while Herbert Morris insists that "A liberal believes that the only legitimate function of criminal law is to prevent private and public harms. This captures liberalism's attachment to the so-called 'Harm Principle.'"[23] Jean Hampton approvingly describes Joel Feinberg's position as being one in which "what makes a state 'liberal' is . . . its rejection of the idea that any enforcement of moral behaviour should include punishment of immoral behaviour which nonetheless has no victim other than the offender himself."[24]

To put one issue to one side: I do not conceive of the debate concerning the harm principle and the criminal law as a debate about whether or not the criminal law should enforce morality, even though it

21. I am grateful to an *Ethics* referee for encouraging me to make this distinction.

22. A. P. Simester and Andreas von Hirsch, *Crimes, Harms and Wrongs* (Oxford: Hart, 2011), 35.

23. Morris, "Professor Murphy on Liberalism and Retributivism," 95.

24. Hampton, "How You Can Be," 114–15.

is sometimes understood in those terms. It is better conceived of as a debate about what subset of morally wrong conduct we can or should criminalize.[25]

In this section, I will introduce four different versions of the harm principle. Each tells us, when criminalizing conduct, to focus on harm,[26] but the instructions of each version of the harm principle differ importantly thereafter. People often talk about the harm principle as if the principle restricts criminal law to harmful (or potentially harmful) conduct, but not all harm principles (including the famous Millian variant) actually say this. Here I want to show how some related, but importantly different, positions are often run together under the banner of "the harm principle." I will then show, in the sections that follow, that none of these harm principles should be endorsed by retributivists.

When people invoke "the harm principle" they (explicitly or implicitly) invoke J. S. Mill. In *On Liberty* Mill famously declares that:

> Mill's Harm Principle (HP_1): "the only purpose for which power can be rightfully exercised over any member of a civilized community, against his will, is to prevent harm to others."[27]

There are two important factors to note about Mill's Harm Principle from a criminalization perspective, which make it importantly different from a principle which restricts criminal law to harmful behavior. The first is that it is not offered as a principle only, or even primarily, for the criminal law. It is offered as a necessary condition of, and the only legitimate purpose of, any exercise of coercive power, which includes all state action and many types of nonstate action as well.[28] The second thing to note is that it focuses not on harmful (or harm-risking) conduct

25. See Joel Feinberg, *Harmless Wrongdoing* (New York: Oxford University Press, 1988), 155; Hampton, "How You Can Be," 106; Gerald Dworkin, "Devlin Was Right: Law and the Enforcement of Morality," *William and Mary Law Review* 40 (1999): 927–46; Ronald Dworkin, "Lord Devlin and the Enforcement of Morals," in *Taking Rights Seriously* (Cambridge, MA: Harvard University Press, 1977), 240–58.

26. Here is an important question about (all versions of) the harm principle: in deciding whether criminalization is justified, does it matter (and, if so, how) why the conduct is harmful? Conduct can be harmful because our states have made it harmful. Neither of the obvious principles—that conduct must simply currently be harmful (regardless of why) or that conduct must be necessarily harmful (and not made harmful by the state)—seem attractive.

27. J. S. Mill, "On Liberty," in *On Liberty in Focus*, ed. John Gray and G. W. Smith (London: Routledge, 1991), 30.

28. In Mill's liberalism, we are to use the harm principle to decide when to intervene using either legal penalties "or the moral coercion of public opinion" ("On Liberty," 30.). However, with regard to nonharmful or self-harming behavior, we are permitted to remonstrate, reason, persuade and entreat (30–31).

but rather the prevention of harm. A third important element of Mill's Harm Principle is that it says that harm prevention is the only purpose for which we may coercively intervene in one another's lives. Therefore it is a version of what Nils Holtug calls "the *justification-based* version of the Harm Principle."[29] That is, harm (or harm prevention) is seen as central to the purpose of interference, not merely a condition on it.

Mill's Harm Principle differs importantly from how many contemporary commentators in criminal law theory discuss the harm principle. First, most assume it is a principle concerning criminalization.[30] Second, most commentators appear to assume that it limits the criminal law to conduct that is either harmful or risks harm to others. As Duff describes the basic message of the harm principle as understood in the philosophy of criminal law, "only conduct that wrongfully harms or threatens to harm others is a suitable candidate for criminalisation."[31]

Duff's harm principle contains two elements—in order to be brought within the criminal law conduct must be both wrongful and harmful (or threaten harm). There are two different ways that we may combine these two elements, wrongfulness and harmfulness, within a theory of the moral limits of the criminal law. In the first, the wrong and the harm must be interlinked. This will give a harm principle like this one:

> Harmful Wrongs Principle (HP_2): the state may criminalize only harmful wrongs—conduct that is both wrong and harmful (or risks harm) to others and is wrong because it is harmful (or risks harm) to others.

Again, this is a "justification-based" version of the harm principle, as harm is not merely a constraint on what we may criminalize but also plays a central role in explaining why we may criminalize the behavior—the crim-

29. Nils Holtug, "The Harm Principle," *Ethical Theory and Moral Practice* 5 (2002): 357–89, 362. Victor Tadros believes that this is the natural way to interpret the harm principle, arguing that "it is in the spirit of the harm principle to see harm not only as a constraint but also as the primary reason that must be offered by states *in favor* of criminalizing some conduct" ("Harm, Sovereignty, and Prohibition," 42).

30. See, e.g., Feinberg's version of the harm principle in *Harm to Others* (New York: Oxford University Press, 1984), 26. As well as restricting the harm principle to criminalization, Feinberg's principle also downgrades harm prevention from being "the only purpose," past "a necessary condition," to being merely "a reason in support" of criminalization. Since Feinberg separates out harm and offense, I think that retributivists should reject Feinberg's ("bold liberalism") position too, since there are some actions which the retributivist ought to regard as punishable or criminalizable which do not offend or harm others. But see *Harmless Wrongdoing*, 323–24, where Feinberg admits that perhaps some (unspecified) harmless wrongdoing may be criminalized.

31. R. A. Duff, "Theories of Criminal Law," in *Stanford Encyclopedia of Philosophy* (2013), http://plato.stanford.edu/entries/criminal-law/.

inal law, on this view, must focus purely on wrongful action that is related to harm. Hampton and Feinberg argue that the liberal state does not punish victimless immoral behavior, which may suggest that the liberal state should only punish behavior that is immoral in virtue of its having a victim (and, furthermore, a victim who is harmed). This is further suggested by Feinberg's claim that liberals see the criminal law as grounded in a "grievance morality."[32] This suggests that not only is the conduct that it is appropriate for the criminal law to prohibit restricted to the harm related, it is its connection to harm that justifies our making it criminal.

Others, however, reject this understanding of the harm principle, and this gives rise to the second way of combining wrongfulness and harmfulness within a theory of criminalization. Simester and von Hirsch, for example, claim that both harmfulness and wrongfulness are necessary conditions of criminalization but that the two need not go together. Wrongs that are wrong independently of harmfulness (harmless wrongs) can be criminalized insofar as the conduct is also harmful or risks harm. Thus Simester and von Hirsch, firm advocates of the harm principle, state that "It is implicit in our analysis that some actions are wrongs prior to, and not in virtue of, any harm . . . that they may cause. Indeed, there is a range of important offences where the wrong is not, in the first instance, grounded in any consequential harm."[33] Under such a view of the harm principle, it is harder to see what justifies the harm principle. The Feinbergian argument from a "grievance morality" seems to justify the harm principle as fundamentally connected to the purpose and justification of the criminal law. Under a harm principle according to which harmless wrongs may be criminalized and punished, but only when they are harmful, it is hard to see, especially from a retributivist perspective, why we would restrict the criminal law to harmful or harm-risking conduct only. If the wrong is a candidate for punishment independently of its harmfulness, why should harmfulness be a necessary condition for criminalization and punishment? I will return to these issues below. For now, we can capture this kind of position with this version of the harm principle:

> Necessary Condition Harm Principle (HP_3): it is a necessary condition of criminalizing conduct that the conduct is harmful (or risks harm[34]) to others.

32. Feinberg, *Harmless Wrongdoing*, 155. See also Hampton, "How You Can Be," 106.

33. Simester and von Hirsch, *Crimes, Harms and Wrongs*, 51.

34. The advocate of HP_3 will owe us an account of what it is to "risk harm" for the purposes of the harm principle (Duff, "Theories of Criminal Law"). At the very least, such an account should rule out "risking harm" as being everything other than conduct that we are certain cannot possibly lead to harm, since the harm principle would then encom-

Whether we accept the Harmful Wrongs or Necessary Condition version of the harm principle, both principles coalesce around what they leave out, the conduct that they demand must not be criminalized: conduct that is nonharmful (and does not risk harm) to others. I will show below that there is some behavior which the retributivist ought to (and, I conjecture, will) agree is properly criminalizable, even though it is either harmless or purely self-harming.

John Gardner and Stephen Shute, however, argue that conduct need not be harmful or risk harm in order to be criminalized under the harm principle.[35] Gardner and Shute make two key moves in specifying their version of the harm principle. The first, which we have already seen, is that they decouple the wrong that justifies punishment from the criterion of harm. Thus, as Gardner says of a would-be harmless wrongdoer (in this case a "harmless rapist"[36]), "This already picks him out as a suitable person to be threatened with punishment (coerced). It is not the job of the harm principle to pick him out again."[37] Gardner and Shute make a second move, developing the harm principle still further away from the intuitive core of harmful wrongs. Not only can harmfulness and wrongness be distinct grounds, they also shift the unit of evaluation for the harm principle (i.e., what it is that needs to be evaluated in terms of its harmfulness) from conduct to legislation. Thus, they make a distinction that Mill himself apparently failed to see. In *On Liberty* Mill argued, as we have seen, that harm prevention must be the aim of interference. This led Mill to think that harmful or harm-risking conduct would be the focus of interference, such that harmless conduct would be left alone: "the principle of liberty requires liberty of tastes and pursuits . . . without impediment from our fellow creatures, so long as what we do does not harm them, even though they should think our conduct foolish, perverse, or wrong."[38] Clearly Mill thought harmless conduct (even if wrongful) to be outside the scope of legitimate coercion.

---

pass possibly all human action, including stepping out of one's front door, switching on the lights, and so on. See Feinberg, *Harm to Others*, 190–93, 216–17. (For HP_2, the internal wrongfulness constraint will take care of the level of risk required.)

35. John Gardner and Stephen Shute, "The Wrongness of Rape," in a collection of Gardner's essays, *Offences and Defences* (Oxford: Oxford University Press, 2007), 1–32. I am grateful to James Edwards, Victor Tadros, and Malcolm Thorburn for discussion of this version of the harm principle.

36. The scare quotes reflect my discomfort with the idea that rape can ever be harmless, and the underlying assumption that harm, by definition, must be consciously experienced. For doubts along these lines, see Tadros, "Harm, Sovereignty and Prohibition," 47–48.

37. Gardner, *Offences and Defences*, 243.

38. Mill, "On Liberty," 33.

In Gardner and Shute's variant of the harm principle, however, harmless conduct can be within the legitimate purview of the criminal law. This is because we do not need to ask whether the conduct is harmful but rather whether criminalizing the conduct will prevent harm: "It is no objection under the harm principle that a harmless action was criminalized, nor even that an action with no tendency to cause harm was criminalized. It is enough to meet the demands of the harm principle that, if the action were not criminalized, *that* would be harmful."[39] This will give us a harm principle like this:

> The Legislation Harm Principle (HP_4): it is a necessary condition of criminalizing conduct that criminalizing the conduct will prevent harm.

I will now show that those who endorse the Retributive Principle ought to reject all four versions of the harm principle.

RETRIBUTIVISM AND MILL'S HARM PRINCIPLE (HP_1)

It is particularly easy to show that retributivists should reject the original, Millian version of the harm principle.[40] The Retributive Principle countenances punishment even when it will not prevent harm. Consider a murderer, who we know is no longer a danger and whose punishment will not prevent harm to others. According to the Retributive Principle, we can still justly punish him, since he deserves it. This requires interfering with him—punishment is a kind of state interference. Mill's Harm Principle, meanwhile, explicitly rules out state (and indeed nonstate) interference unless doing so prevents harm. Indeed, even if the punishment of the murderer were to prevent some harm, for Mill this is the only reason for which punishment could be pursued, and so considerations of desert could not play a role in justifying interfering with the individual. HP_1 and the Retributive Principle, therefore, are squarely at odds, both over whether we can interfere when this will not prevent harm, and over whether harm prevention is the only legitimate purpose of intervention.

RETRIBUTIVISM, THE HARMFUL WRONGS PRINCIPLE (HP_2), AND THE NECESSARY CONDITION HARM PRINCIPLE (HP_3)

It is not so straightforward, however, for me to show that retributivists should reject the Harmful Wrongs or the Necessary Condition versions of

39. Gardner and Shute, "The Wrongness of Rape," 29, and see further 29–32.

40. The tension between backward-looking theories of punishment and forward-looking liberal concerns is explored in Thaddeus Metz, "How to Reconcile Liberal Politics with Retributive Punishment," *Oxford Journal of Legal Studies* 27 (2007): 683–705.

the harm principle. This is because, unlike HP_1, they are logically compatible with the Retributive Principle. Endorsing the conjunction of the Retributive Principle and the Harmful Wrongs Principle simply requires the retributivist to endorse one of the following positions:

- (P_1) Only conduct which deserves (or will deserve[41]) punishment may be criminalized, and only Harmful Wrongs (i.e., conduct that is wrongful because it is harmful or risks harm) deserve punishment.
- (P_2) Only conduct which deserves (or will deserve) punishment may be criminalized, only wrongful conduct deserves punishment, and Harmful Wrongs exhaust wrongfulness.
- (P_3) Only Harmful Wrongs may be criminalized, and only criminalized conduct deserves punishment.

(P_1) and (P_2) might be endorsed by the moral retributivist, while the legal retributivist might endorse (P_3). In order to get the retributivist to reject (P_1)–(P_3), we need to show her a class of conduct that is wrongful and criminalizable, but that is not wrongful in virtue of its harmfulness.

Endorsing the conjunction of the Retributive Principle and the Necessary Condition Harm Principle requires the retributivist to endorse one of the following:

- (P_4) Only conduct which deserves (or will deserve) punishment may be criminalized, and only harmful (or harm-risking) behavior can deserve punishment.
- (P_5) Only conduct which both deserves (or will deserve) punishment and is harmful (or risks harm) may be criminalized.[42]
- (P_6) Only harmful or harm-risking behavior may be criminalized, and only criminalized conduct deserves punishment.

In order to get the retributivist to reject (P_4)–(P_6) we must show her a class of conduct which she believes deserves punishment, or ought to be criminalized, even though it is harmless or only self-harming.

Consider, now, the body of law which I will call "self-protecting criminal law." This body of law uses the criminal law to protect, and respond to violations of, the criminal justice process itself. Examples of such laws include those that criminalize bribing witnesses, escaping mandated pun-

41. This parenthetical clause allows for conduct that is not morally deserving of punishment in the absence of a criminal law prohibiting it but becomes so deserving once criminalized (e.g., driving on the "wrong" side of the road).

42. This is the position taken in Douglas Husak, *Overcriminalization* (New York: Oxford University Press, 2008), chap. 2. Although Husak would most likely reject the Retributive Principle (chaps. 3–4), he accepts comparative proportionality and so will be brought within the ambit of the thesis below.

ishment, assisting someone to escape mandated punishment and perjury in the criminal trial. The rationale for all of these laws seems to be the protection of the criminal justice process—of ensuring just conviction and punishment. Some conduct proscribed by self-protecting criminal law is not directly harmful: no one is necessarily directly harmed if I accept a bribe and then fabricate testimony which helps a criminal escape punishment.

To help fix ideas, consider the murderer, whose punishment is justified even if it won't prevent harm. So far, this is consistent with both HP_2 and HP_3, since in murdering someone, they harmed someone and murder is wrongful because it is harmful.[43] Now, imagine that after he is convicted the guards who must transport him to prison pull over to the side of the road, and open the back doors of the van. Off the murderer runs, into the night, never to be seen again. Should such conduct be criminalized and punished? Surely. But the retributivist who accepts this should deny (P_1)–(P_6) and thus should reject HP_2 and HP_3. First, consider HP_2. In order to support HP_2, the retributivist needs to endorse one of (P_1)–(P_3), which each state that only wrongs that are wrong in virtue of their harmfulness (or the risk of harm) may be criminalized and punished. Is letting the murder go wrong because it is harmful? No. For a start, it may not even be harmful (a point to which I will return below), but even if it were harmful, it surely is not wrong in virtue of its harmfulness. The reason the drivers must not let the man out of the van is because if they do so, he will escape his just punishment—they interfere with the criminal justice process. Letting anyone out of the van may risk harm, but the distinctive wrongfulness of letting this particular person out is to be found in helping them escape their just punishment.[44] Now, if the murderer's punishment were justified by the prevention of future harm (through incapacitation, special deterrence or general deterrence) then the drivers would be undermining a harm-preventing institution, and the wrong would be one of interfering with a harm-preventing institution. In such a case, then the advocate of HP_2 could legitimately claim that the wrong

43. It is, of course, philosophically controversial whether death, and therefore murder, is harmful (Steven Luper, "Death," in *Stanford Encyclopedia of Philosophy* [2009], http://plato.stanford.edu/entries/death/). I set these complications aside, since, in political and legal philosophy, I have never seen anyone claim that murder is outside the parameters of the harm principle. Whatever people mean to capture by the harm principle, they mean for it to include murder.

44. An editor of *Ethics* suggests that in breaking their contracts, the van drivers harm their employer. But it is far from clear that merely breaking a contract is necessarily harmful—promise breaking needn't be harmful, and we normally keep employment contracts outside the criminal law. For those convinced by the editor's claim, however, we can simply change the example, such that the prisoner is released by people who haven't signed any contract promising not to do otherwise.

is harmful wrong—that is, interfering with a harm-preventing institution is wrongful because it is (indirectly) harmful.

However, the retributivist cannot legitimately claim that the wrong here is of that sort. That is because the retributivist sees the criminal justice process as being about giving people their just deserts, or reinstituting a fair division of benefits and burdens between the law-abiding and the criminal—in other words, she sees the criminal justice process (at least in this case) as serving the value of retributive justice. Since the murderer was to be punished whether this prevented harm or not, then she cannot see the wrongness of interfering with the criminal justice process as a harmful wrong, however indirect the harm. If the value of retributive justice is important enough for us to punish people on that basis, isn't the wrongfulness of interfering with it the kind of wrongfulness that itself can attract criminal sanction? If this is the case, we should reject (P_1)–(P_3) and therefore HP_2.

Retributivist supporters of HP_3, however, do not need to see the wrong of interfering with the criminal justice process or just punishment as a harmful wrong. Therefore, they can agree that the wrongfulness of letting the murderer go is to be found in disrupting the delivery of retributive justice. (P_4)–(P_6) (the propositions necessitated by the conjunction of HP_3 and the Retributive Principle) simply require that, in order to be criminalized, the conduct also be harmful or risk harm. But what should we then say of cases where people deserve punishment, but where punishing them is not preventing harm and where (we can stipulate[45]) we know that punishing them is not preventing further harm? Imagine our murderer is just such a person—he is an aged criminal (though still deserving of punishment), no threat to anybody, and his punishment will not serve as a deterrent. Releasing him, therefore, does not risk harm to anyone. If we were instrumentalists about punishment, requiring punishment to have further good effects, then we would not be able to jus-

45. This stipulation is somewhat unrealistic, as the drivers most likely cannot be certain that their conduct does not, at least, risk harm. First, I will defend the unrealistic stipulation. Since the harm principle is proposed as a fundamental principle of political morality, about the proper limits of the criminal law, it should be able to cope with such a stipulation—it is only by thinking through cases in which some conduct is clearly harmless that we can assess whether harm plays any role in the legitimate criminalization of that conduct. Even if we never, or rarely, come across a clearly harmless case, we can ask whether it can, in principle, be criminalized (Ripstein, "Beyond the Harm Principle," 223). Indeed, it would be especially odd for retributivists to complain about the use of such examples, since such examples (where we know that punishment won't prevent harm) are central to developing the distinctiveness and plausibility of retributivist theory. Second, we can, however, relax the stipulation. As I observed above (n. 34), a plausible harm principle will not allow the criminalization of insubstantial risks. Let us imagine, instead, that the drivers have no good reason to believe that they create (and do not create) any greater risk of harm than I do when I let my ninety-two-year-old grandfather out of the car at his bridge club. I take it my conduct will be outside HP_3 and thus so will the drivers'. I am grateful to an *Ethics* editor for comments here.

tify punishment in this case, and we ought to release him. But if we are retributivists, as per the Retributive Principle, then we think the murderer's punishment justified, even though this does not prevent further harm. Given that he can justifiably be punished, are we not permitted—on that basis alone[46]—to criminalize the prevention of, or the attempted prevention of, him receiving his just punishment, and punish those who break such norms? If we believe that the criminalization and punishment of the drivers is justified, then we must reject (P_4)–(P_6), for we have found some conduct that ought to be criminalized even when it is harmless.

I want to emphasize here that my example starts from a harm principle-approved crime (murder). Therefore, I am not seeking to show the harm-focused retributivist a completely independent crime or set of crimes that she ought to accept. I am, instead, trying to show her a set of laws which protect and respond to violations of the (retributive and harm-focused) criminal justice processes that she already accepts.[47] The retributivist cannot deny that murder should be criminal or that retributive justice is a value. The processes which self-protecting criminal laws protect draw their value from retributive justice and retributive justice does not draw its value from preventing harm. If we can criminalize violations of retributive justice in and of themselves, then we can criminalize behavior that doesn't harm or risk harm to others.

If the above arguments go through, the retributivist who wants to hang on to the harm principle might declare herself willing to bite the bullet, and to argue that retributive justice, in and of itself, is not worthy of the protection of the criminal law, and that attempts to interfere with or to frustrate it are not, in the absence of harm or the risk of harm, worthy of criminalization or punishment: the drivers' conduct must not be criminalized or punished.

46. Some will object that even when conduct is harmless, the fact that the criminal law is an inevitably "blunt instrument" means that the harm principle allows for regulation "in terms of average cases"—for example, speed limits (Simester and von Hirsh, *Crimes, Harms and Wrongs*, 45–46). To apply this thinking to the present case would essentially see the criminalization of the harmless release of the prisoner as an unfortunate side-effect of the criminal law's blunt nature and would be to deny that the criminalization of the drivers' conduct is in and of itself warranted. For reasons articulated below, I do not think that retributivists will, or should, accept this.

47. This element of the argument (which is, I believe, what makes it so powerful when directed at the retributivist) suggests a way that the harm-focused retributivist could amend the harm principle's scope—by restricting it to initial offenses (like murder). Such a position would nevertheless represent an important downgrading of the harm principle, and it would explicitly allow the criminalization and punishment of conduct that is beneficial to some and harmful to none. However, an argument would need to be made for this restriction—in essence, we'd need to know why retributive justice is the only impersonal value that can warrant criminalization. Although he does not cover the cases I am interested in here, Metz's "expressive liberalism" would be an interesting avenue to explore here ("How to Reconcile," 698–703). I am grateful to an *Ethics* editor for comments here.

I think this would be an odd position for the retributivist to adopt, and it is worth noting that in presenting and discussing this work and these cases I have not found anyone prepared to say that releasing the murderer should not be criminalized. Rather, retributivists tend to look for someone who is being harmed (on which, see below), or amend the harm principle (often in the direction of HP_4—on which, again, see below). What is wrong with saying that violations or frustrations of retributive justice ought not, in and of themselves, to be criminalized? First, it seems implausible, as a matter of policy, that we should have to show potential harmfulness before criminalizing perjury, bribing witnesses, absconding, or assisting prison breaks. If we think that punishment is justified by retributive justice, these actions are wrong because they are an affront to retributive justice and our attempts to secure it, and are criminalizable on that basis.

Second, it reveals a kind of inconsistency, though not the strict logical inconsistency that occurs when the retributivist endorses HP_1. To refuse to allow the criminal law to protect the delivery of retributive justice seems to require an unstable bifurcation in our attitude toward retributive justice. It requires seeing retributive justice, on the one hand, as the value that our criminal justice system must promote through punishment, but, on the other, as the kind of value that is not the sort of thing we should issue threats to protect, or punish people for frustrating our attempts to promote—the sort of thing the criminal law should see as none of its business. Since the retributivist has already accepted that we punish in order to further retributive justice, wouldn't it be odd to say that it would be inappropriate to use that system to criminalize and punish people for their attempts to thwart its promotion? Since, for the retributivist, all punishment is based on achieving retributive justice, it would seem strange to deny that we can punish people for frustrating the promotion of that very value. For these reasons, I think that denying that self-protecting criminal law is, in and of itself, justified is an implausible position, both politically and morally, for the retributivist to adopt.

RETRIBUTIVE JUSTICE AND IMPERSONAL VALUE

Retributive justice is plausibly regarded as impersonally valuable, or impersonally morally important (these two formulations apply to telic and deontic retributivism, respectively).[48] That is, the value or importance of securing retributive justice is not grounded in its being good for some

48. Indeed, retributive justice is so identified with the concept of impersonal value that it is used an example of an impersonal value that many of us accept in order to bolster the thought that others (like distributive equality) might exist. See Larry S. Temkin, "Equality, Priority and the Levelling Down Objection," in *The Ideal of Equality*, ed. Matthew Clayton and Andrew Williams (Basingstoke: Palgrave Macmillan, 2002), 126–61.

particular person. Return to our murderer. Plausibly, punishing him is bad for him. It is also plausible to believe that it is not necessarily good for anyone in particular that he is punished. Therefore, the value or importance of punishing him is impersonal. The wrong, and the criminalizable wrong (if we are prepared to accept, as all retributivists surely will, that such conduct is criminalizable) of releasing him is to be found in the offense to the impersonal value or importance of punishing murderers—the value or importance of doing retributive justice. HP_2 and HP_3 rule out criminalization purely on the basis of impersonal value or importance—they require conduct to be (potentially) bad for someone (i.e., harmful) before we countenance criminalization. Accepting that retributive justice is such a value, and that it warrants the protection of the criminal law, therefore means rejecting HP_2 and HP_3. But, beyond that, it also means accepting that actions which are good for somebody (i.e., the criminal) and bad for nobody—in other words, Pareto improvements—can nevertheless be criminalized.

If retributive justice is viewed in this way, then what I have shown here is not only that retributivists should reject the harm principle but also that they should reject all attempts to understand the criminal law through the lens of person-affecting or "grievance" morality.[49] Unlike examples which try to show that the harm principle is faulty by showing other ways we might negatively affect someone without harming them, this class of cases shows that retributivists ought to accept the criminalization of behavior that benefits someone, and is bad for nobody. Therefore, if the retributivist chooses the impersonal value or importance of retributive justice over the harm principle, her rejection of the harm principle must be more radical than that often recommended—it is not harm but rather the idea of negatively affecting other people at all that she must reject as a necessary condition on criminalization.

IS INTERFERENCE WITH RETRIBUTIVE JUSTICE HARMLESS?

Perhaps retributive justice is not best thought of as an impersonal value. Seeing retributive justice as impersonally valuable means seeing thwarting retributive justice as a victimless crime. Some will deny that retributive justice is impersonally valuable and will argue that I have ignored some potential beneficiaries of retributive justice and thus victims of behavior that frustrates or disrupts it. These potential victims include the

49. Of course, the state will have a grievance, since if it is justifiably punishing, it must have the right to punish, and so its right is violated when criminal justice is interfered with. But, crucially, the state is not a person, and it does not, for the retributivist, have this right in order to protect its citizens from harm or other bad effects but rather in order to pursue retributive justice. Therefore, the state's grievance is based on the protection of impersonal value, and so is outside the realm of interpersonal grievance morality. I am grateful to Zofia Stemplowska for discussion here.

victim of the original (harmful) crime, who is denied "justice"; the original offender, who does not experience the appropriate response to their crime; and the wider community who do not get to live in a (retributively) just society. If any of these people is negatively affected such that they are harmed by releasing a criminal deserving of punishment, then we will be able to say that releasing the murderer is harmful conduct and thus, at the least, within the scope of HP_3. However, note that this would not save the combination of HP_3 and the Retributive Principle. Rather, it requires rejecting or amending the Retributive Principle. This is because the Retributive Principle countenances punishment even where doing so does not prevent harm. Each of the suggested responses above, however, claims that deserved punishment, conceptually, must prevent harm, since a wrongdoer not getting their deserved punishment is necessarily harmful to someone, and so deserved punishment that doesn't prevent harm is a conceptual impossibility.

Nevertheless, those who support such alternative understandings of the value of retributive justice (and thus suitable amendments to the Retributive Principle) should still not combine them with HP_2 or HP_3. With regard to the original victim, it is worth remarking that since the state prosecutes criminal offenses, it is not clear that we think that "justice" is good for victims, or, at least, we do not pursue retributive justice on that basis. Therefore, there may be cases where justice is not good for a victim, even if justice is, in general, good for victims (and murder may be a good example of this). In addition, it is important to note that the definition of harm will have to travel some distance for "other people being punished in my name" to be considered the kind of interest or source of well-being whose denial constitutes "harm." However, even if the original victim is ordinarily harmed when justice is not done or is frustrated, we can point to cases where it is the original victim who colludes with the original offender to get them off, perhaps by accepting a bribe. It is clear that in such cases the original victim judges that they are made better off by having the cash and no justice, rather than justice and no bribery money. But we could insist that they are wrong, and that they are harmed by not getting justice. Even if they are harming themselves, however, they are harming *themselves*, and HP_2 and HP_3 do not allow us to criminalize purely self-harming action.

The same point applies to the original offender. Personally, I find it wildly implausible that we harm someone by not punishing them (when such punishment is deserved). This seems to duck the real normative difficulty of punishment—how can we justify harming people in this way?—by simply reinventing just punishment as a source of personal good. But let us grant, arguendo, the claim that deserved punishment is good for the punished, such that they are harmed if such punishment is disrupted. This is still not enough to save the harm principle for the retributivist, for

if the convicted offender is a willing participant in plans to undermine retributive justice, or if they act alone and, for example, escape from jail, then even if being on the receiving end of retributive justice is good for us, the offender consents to such harm, or inflicts it on himself. Since HP_2 and HP_3 disallow the criminalization of merely self-harming behavior, they would disallow criminalizing offender-led jail breaks when these are not harmful to others. Once again, if retributivism is to be retained, the rejection of the harm principle would need be more radical than the usual attempts to go beyond it. Here, in order to criminalize such behavior, we'd need to criminalize purely self-harming behavior.

Other potential victims are the wider society whose opportunity to live in a society where justice is done is set back.[50] How could somebody else escaping justice harm me? Here are two possible stories about why we are harmed when retributive justice is set back. On the first, we want to punish wrongdoers, and so those wrongdoers who escape justice frustrate this aim. In order for interference with the achievement of this desire to engage the harm principle, however, we would have to (*a*) hold a desire-fulfillment or achievement view of welfare; and (*b*) hold that setbacks in welfare constitute harm in the relevant sense. Endorsing both of these positions, however, would reduce the harm principle to a meaningless restriction on the scope of the criminal law. All that would be required in order to make conduct within the ambit of the harm principle and thus potentially[51] criminalizable is that someone must have taken it as their aim to stop such conduct or to live in a world without such conduct. Thus, provided someone wants to live in a world without homosexual sex (or wants to protect an impersonal value), then homosexual sex (or offense to that impersonal value) will be within the ambit of HP_3—we would have license to make anything harmful, simply by giving ourselves goals that conflict with such behavior.

Another way that we may be harmed when somebody escapes just punishment is that in doing so they perform or create an injustice, and injustice is inherently bad for us and harms us.[52] Perhaps injustice is always bad for us—justice is a valuable goal and it is plausible to believe that my life goes, in some way, better when I live in a just society. But what the advocate of this response requires to be true is not only that injustice is bad for us, but that retributive injustice harms us, and, furthermore, harms us in the sense invoked in the harm principle. It is this last claim

50. I am grateful to Antony Duff, Catriona McKinnon, Zofia Stemplowska, and Victor Tadros for comments and discussion here.

51. The wrongfulness criterion must be remembered here.

52. For such a view, see John Broome, *Weighing Goods* (Oxford: Oxford University Press, 1991), 168–69. For criticism, see Temkin, "Equality, Priority and the Levelling Down Objection," 147–51.

that I think is particularly problematic for the retributivist who advocates the harm principle. While I have tried thus far to discuss the harm principle while remaining as neutral as possible on the correct understanding of harm,[53] it is clear that each time we stretch the concept away from its intuitive core, we render the harm principle a little less distinctive as a claim. To say that injustice automatically harms, however, stretches things such that the harm principle becomes completely worthless—it saves the combination of retributivism and the harm principle at the cost of reducing the harm principle's role in criminalization decisions to nothing, rendering it nothing more than a platitude.

Let us call the intuitive core of harm, where I am directly affected by some conduct, like when my leg is cut off, "well-being harm" (I won't try to say here exactly what is and isn't in this category). Now, let us call the harm inherently caused to me by someone else doing or causing a retributive injustice (or undermining a just institution or a fair distribution), whether or not it directly affects anyone, "justice harm." Were HP_2 and HP_3 to include justice harms, as well as well-being harms, within their scope, they would become redundant. The harm principle is supposed to be a substantive claim about the proper scope of the criminal law and a tool with which to assess and critique our positive law. But if some conduct's harmfulness can depend on its either being part of the proper scope of the criminal law, or its being part of the positive criminal law, then the concept of harm cannot help us determine the proper scope of the criminal law and/or assess our positive criminal law.

Take, first, the moral retributivist. She believes that certain kinds of moral wrongdoing ought to be punished. If she advocates the HP_2 or HP_3, then she believes that only wrongful conduct which harms (or risks harm) ought to be punished. However, if retributive injustice harms, then all conduct that is punishable harms, since all punishable conduct creates an injustice or an unfair distribution (for, at least until the actor is punished, there is a retributivist injustice or an unfair distribution) and (provided there is some risk of not being caught) undermines just institutions. If all that ought to be punished harms, then the harm principle is true by definition: only harmful conduct is punishable, because punishable conduct is by definition harmful.

53. Harm is a notoriously slippery concept. See Holtug, "The Harm Principle"; Davis, "Harm and Retribution"; Bernard E. Harcourt, "The Collapse of the Harm Principle," *Journal of Criminal Law and Criminology* 90 (1999): 109–94. For recent work on the concept of harm, see Seana Valentine Shiffrin, "Wrongful Life, Procreative Responsibility, and the Significance of Harm," *Legal Theory* 5 (1999): 117–48; Matthew Hanser, "The Metaphysics of Harm," *Philosophy and Phenomenological Research* 77 (2008): 421–50; Judith Jarvis Thomson, "More on the Metaphysics of Harm," *Philosophy and Phenomenological Research* 82 (2011): 436–58; Matthew Hanser, "Still More on the Metaphysics of Harm," *Philosophy and Phenomenological Research* 82 (2011): 459–69.

Now take the legal retributivist. She believes that legal wrongdoing ought to be punished. If she advocates the harm principle, then she also thinks that we should decriminalize nonharmful or self-harming conduct and should only criminalize conduct that harms others. However, if retributive injustice harms, then all legally defined wrongdoing harms, since all legal wrongdoing creates a retributive injustice. If all legal wrongdoing is harmful, then the harm principle cannot be used to assess or critique our positive criminal law, for all conduct proscribed by criminal law will, by definition, already be harmful. And any conduct will be made harmful by criminalizing it. Only if harm has a nonlegal definition can it be used to assess, amend, or guide us in enacting our criminal law.

The harm principle was supposed to tell us something about what conduct is punishable or ought to be within the criminal law. But if punishable or criminalized conduct is automatically harmful (as it is when we include justice harms within the concept of harm), then the harm principle cannot tell us anything about what should be criminal or is punishable—it is no longer a substantive claim.

RETRIBUTIVISM AND THE LEGISLATION HARM PRINCIPLE (HP_4)

In order to reject, on behalf of the retributivist, HP_2 and HP_3, we have, in part, looked at cases in which harmless (or, perhaps, self-harming) conduct nevertheless ought to be criminalized. However, recall that John Gardner and Stephen Shute recommend this version of the harm principle, which does not require conduct to be harmful or risk harm:

> The Legislation Harm Principle (HP_4): it is a necessary condition of criminalizing conduct that criminalizing the conduct will prevent harm.

What does our argument here tell us about this version of the harm principle? Since I have pointed to conduct that could be harmless, but that the retributivist ought to think criminalizable, a harm principle that allows the criminalization of harmless conduct is obviously to be preferred (if one is a retributivist). So, the present article may, for some, be an argument for retributivists accepting HP_4. And, indeed, retributivism is logically compatible with HP_4. The conjunction can be endorsed if the retributivist endorses one of the following positions:

(P7) We may criminalize conduct only if criminalizing the conduct will prevent harm. Only criminalized conduct deserves punishment.

(P8) A necessary condition of criminalizing conduct is that it will prevent harm. A separate necessary condition is that the conduct deserves (or will deserve) punishment.

However, I think that the retributivist should reject HP_4 as well. Consider our cases of harmlessly assisting people escape justice. Even on the HP_4 version of the harm principle, this may be legitimately criminalized only if doing so will prevent harm. (Indeed, on HP_4, even harmful behavior, like assault and murder, can only be criminalized if doing so will prevent harm.) From the retributivist perspective, should we criminalize harmless interference with the pursuit of retributive justice (or assault and murder) only if doing so prevents harm? Intuitively, this does not seem right. Surely the conduct is worthy of criminalization and deserving of punishment in and of itself.

The retributivist, however, should not accept HP_4 principally because she should not endorse its grounds. What is the justification for the harm principle as understood by Gardner and Shute? It does not demand that we punish for harmful or harm-risking behavior, and it does not try to build an understanding of criminal law founded on harmful wrongs. Therefore, we can ask, if the harmless wrongdoer is a suitable target for the criminal law in the absence of harm, why is there a harm-prevention requirement on criminalizing his conduct? An answer to this is provided by Gardner and James Edwards. They write: "Wherever [the criminal law] goes, it spreads its own harms, not only the intentionally inflicted harms of punishment but many harmful side-effects too. The harm principle requires that it do so only in return for harms avoided. To defend the harm principle as a strict limit on criminalization is thus to defend the proposition that harm takes lexical priority over other ill-effects of wrongdoing, since no amount of harmless pain or inconvenience can ever warrant an infliction of even very slight harm."[54] Therefore, Gardner and Edwards support HP_4 because they buy this claim:

(P9) Harm is lexically dominant. Harm may be done only if it prevents harm.

Should we accept (P9)? Personally, I doubt it. But retributivists certainly should not. (P9) says we can only harm if it prevents harm. Retributivists could deny (implausibly, I believe) that punishment harms. This means that they could accept (P9), but since they would not think just punishment harm, (P9) would not restrict the use of this tool. If, however, the retributivist believes (like Gardner and Edwards) that punishment is harm, then they have already rejected the key proposition, (P9). They believe that pursuing retributive justice can outweigh the importance of avoiding harm: the harms of punishment can be imposed in order to

54. John Gardner and James Edwards, "Criminal Law," in *The International Encyclopedia of Ethics*, ed. Hugh LaFollette (Oxford: Blackwell, 2013), 1190. See also Husak, *Overcriminalization*, 203–5.

achieve retributive justice. Given this, they cannot endorse the claim that avoiding harm takes lexical priority over everything else.

The retributivist might accept the following:

(P10) Undeserved harms are lexically dominant. Undeserved harm may be done only if it prevents undeserved harm.

(P10) places a firm wedge between harms that genuinely further retributive justice and those that do not. The undeserved harms caused by the criminal justice process will take the form of unjust punishment and side-effect harms (such as the harms done to the friends and relatives of criminals). If the retributivist accepts (P10), then two things follow. First, the retributivist vastly reduces the role that her retributivism can play in any actual punitive scheme. The retributivist can allow for just punishment (deserved harm), when no undeserved harm will be brought about, without preventing undeserved harm. However, since all actual punitive schemes produce undeserved harms, through side-effects and errors, in setting up such any actual scheme we must focus first and foremost on preventing undeserved harms. If punishing our murderer causes or risks, but does not prevent, undeserved harm, then we must not do it—no amount of retributive justice makes up for a speck of undeserved harm. Therefore, accepting the lexical priority of preventing undeserved harm means rejecting the Retributive Principle in a world like ours, where we cannot be sure that only deserved harms will be done.

Second, accepting the lexical priority of preventing undeserved harms over other goals means accepting the harm principle not as a fundamental principle about what kinds of conduct the criminal law ought, in principle, to be concerned with but rather as a contingent principle. It is a principle that we endorse only when and because the criminal law makes mistakes or harms those it does not intend to—it is the risk of undeserved (and thus unintended) harm that drives the restriction. We can reject HP_4 and criminalize conduct even when doing so will not prevent undeserved harm, if we know that only deserved harms will result. But since our institutions cannot do this, then we should accept HP_4 in the real world.

Put together, these two points allow us to endorse both the harm principle and retributivism but not to put them together within a system of criminal law (we cannot endorse them together). That is because the two principles operate at different levels. Retributivism, on this view, is an ideal principle, which we can use only where retributive justice will be delivered with perfect precision and will not harm anyone undeservedly. The harm principle, meanwhile, is not an ideal principle—it is a principle only for use in worlds (like ours) where criminal justice is delivered

imperfectly, where undeserved harm might be done by the system. In other words, the harm principle is a principle for the kind of world where retributive justice will not get a look in.

COMPARATIVE PROPORTIONALITY AND THE HARM PRINCIPLE

Thus far I have shown that those who endorse the Retributive Principle should not accept any version of the harm principle. Even retributivists, however, who believe that punishment must be justified with reference to preventive concerns (and thus reject the Retributive Principle) could fall within the ambit of the thesis defended here, provided they accept that a principle of comparative proportionality should play a role in sentencing that is not fully subservient to preventive concerns.[55] This is important, since it brings many more theories and theorists within the ambit of my thesis. Indeed, some have argued that a commitment to proportionality is essential to retributivism,[56] and Matt Matravers argues that the "retributivist" resurgence of the second half of the twentieth century was really a demand for proportionality in sentencing, rather than a demand that punishments be deserved.[57]

Comparative proportionality demands two things—that like cases be treated alike, and that unlike cases be treated differently in proportion to their differences.[58] Theories which give a fulsome role to comparative proportionality in punishment fall within the ambit of my thesis because comparative proportionality can recommend nonbeneficial or nonharm preventing punishment and therefore, like retributive justice, can be seen as impersonally valuable or important. To see this, consider a society which has only two crimes: *A*-ing and *B*-ing. *B*-ing is twice as bad as *A*-ing, such that comparative proportionality demands that *B*-ing be punished twice as much as *A*-ing. Imagine that, as things currently stand, *A*-ing is punished with two years' imprisonment, and *B*-ing with four, in accordance with comparative proportionality. The legislature learns that

55. This will include those who see retributivism/desert as having a role only in answering the second and third of Hart's questions (whom may we punish? and how much?) and not in answering his first (why may we punish?). See H. L. A Hart, *Punishment and Responsibility* (Oxford: Oxford University Press, 2008), 8–13.

56. Andrew von Hirsch, "Proportionate Sentencing: A Desert Perspective," in *Principled Sentencing*, 3rd ed., ed. Andrew von Hirsch, Andrew Ashworth, and Julian Roberts (Oxford: Hart, 2009), 115–25 at 117; Douglas Husak, "Retributivism, Proportionality, and the Challenge of the Drug Court Movement," in *Retributivism Has a Past: Has It a Future?*, ed. Michael Tonry (New York: Oxford University Press, 2012), 214–33 at 214.

57. Matt Matravers, "Is Twenty-First Century Punishment Post-Desert?," in Tonry, *Retributivism Has a Past: Has It a Future?*, 30–45.

58. Comparative proportionality is sometimes referred to as "ordinal proportionality," and can be contrasted with limiting proportionality, which simply demands that punishments for a given crime do not exceed some absolute level.

if it increases the punishment of *A*-ing by one year, to three years, this will dramatically decrease incidences of *A*-ing. This move will, in and of itself, be just—the decrease in *A*-ing will justify the increased sentence on the preventively oriented punitive theory endorsed by the society. Comparative proportionality, meanwhile, demands that the sentence for *B*-ing be increased from four to six years, so that the punishment for *B*-ing remains twice that of *A*-ing.[59] However, *B*-ers are more determined than *A*-ers and less easily deterred. The legislature learns that incidences of *B*-ing will not decrease at all by increasing the sentence from four to six years. Thus the increase in punishment for *B*-ing, introduced in the name of proportionality, serves no preventive function whatsoever—a sentencing regime of three years for *A*-ing and four years for *B*-ing would be just as effective preventively, but it would be comparatively disproportionate and thus disallowed.

Given this, once the increases—from two to three and four to six years, respectively—are enacted, *B*-ers serving the last two years of their sentence are being punished purely to serve the goal of proportionality. Plausibly, they are made worse off, while nobody is made any better off, and if so comparative proportionality (if it has value) has impersonal value.

Since punishing people interferes with them, and this interference is not being done with the aim of preventing harm, then HP_1 is to be rejected. Since it requires harming people without preventing harm, then those who endorse it should reject the lexical dominance of harm prevention and thus reject the grounds of HP_4. And, if we can criminalize people interfering with the last two years of the *B*-ers' sentences, then we can criminalize harmless conduct and thus should reject HP_2 and HP_3.

Imagine a group who starts to break *B*-ers out of jail only in the last two years of their sentence. They develop a harmless technique of jailbreaking, and publicly explain that they will not break any *B*-ers out until they have served four years, so all *B*-ers will know that they will serve four years. Since this sentence has the same preventive effect as six years, the group do no harm to anyone else through weakening the primarily preventive enterprise of punishment. Therefore, HP_2 and HP_3 would demand that their conduct be allowed by criminal law, since it is harmless. And yet, if we believe in a principle of comparative proportionality, do we believe that a society cannot legitimately legislate to protect proportionate punishments? If comparatively proportionality is so important that it explains why we should punish *B*-ers for an extra two years in the first place,

59. I ask you to imagine that the overall severity of a punishment is directly proportionate to the length of a sentence, such that you always punish twice as much by punishing twice as long. This is obviously controversial.

is it not also worth protecting with criminal law? If so, HP_2 and HP_3 must be rejected, and impersonal values, and not only harms or harm-like concepts, can justify criminalization.[60]

CONCLUSION

In this article I have argued that two commonly held positions in the philosophy of criminal law—that punishment is based on desert (or that punishments must be comparatively proportionate) and the harm principle—should not be endorsed alongside one another. I have not tried to say whether we should reject the harm principle or retributivism, but I have shown four different versions of the harm principle should not be endorsed alongside retributivism, and so one of them must be rejected. I think retributive justice (and comparative proportionality) are best seen as impersonal values—if they are good, they are not good because they are good for anyone. And if we believe that an impersonal value is at the very center of the justification of punishment, that we can punish in order to serve retributive justice or proportionality, then it seems odd to think that we cannot criminalize and punish in order to protect or respond to violations or disruptions of that very value. And if we can criminalize on the basis of impersonal values, then the criminal law is not only properly concerned with behavior that negatively affects other people or beings.

60. Obviously the retributivist may insist that proportionality is good for someone. My replies to such a claim will be along the lines (mutatis mutandis) of those made in the section "Is Interference with Retributive Justice Harmless?"