Legal Theory, **20** (2014), 253–285.

doi:10.1017/S135232521500004X

HARM PRINCIPLES

James Edwards*
Worcester College, Oxford

ABSTRACT
Much time has been spent arguing about the soundness of "the harm principle." But in the philosophical literature there is no single such principle; there are many harm principles. And many objections pressed against "the harm principle" are objections to only some of these principles. The first half of this paper draws a number of distinctions between harm principles. It then argues that each harm principle is compatible with many other principles that impose limits on the law, including but not limited to other harm principles. The second half of the paper applies the lessons of the first to a number of prominent objections to "the harm principle." That principle has been accused of a) being underinclusive; b) misrepresenting the reasons why many act-types ought to be legally proscribed; c) permitting lawmakers to treat people as mere means of achieving their ends; and d) being overinclusive. The paper argues that one harm principle survives all four objections.

Many philosophers and legal theorists have spent time arguing for or against "the harm principle." Their arguments can be evaluated only once we know what that principle *is*. But the literature in question contains no single harm principle; it contains many harm *principles*. And many objections pressed against "the harm principle" are objections to only some of these principles. To make progress with the question of whether any harm principle is sound, we must thus distinguish between the different harm principles on offer. We must also remember that each harm principle is compatible with many *other* principles: in addition to endorsing a harm principle, one may consistently endorse other principles that impose limits on the law,

*I am grateful to audiences at the Political Philosophy Workshop in Cambridge, the Jurisprudence Discussion Group in Oxford, and the Legal Philosophy Workshop in Pennsylvania. Special thanks to Alex Guerrero, who was my commentator in Pennsylvania, and to Luís Duarte d'Almeida, Matthew Hanser, Doug Husak, and Andrew Simester for written comments.

including but not limited to other harm principles. The first half of this paper seeks to vindicate these claims. Sections I to IV distinguish between different candidate harm principles. Section V argues that each candidate can be endorsed alongside many other principles.

The second half of the paper applies the lessons of the first to a number of prominent objections to "the harm principle." That principle has been accused of a) being underinclusive; b) misrepresenting the reasons why many act-types ought to be legally proscribed; c) permitting lawmakers to treat people as mere means of achieving their ends; and d) being overinclusive. Sections VII to X argue that one harm principle survives all four objections. This does not, of course, show that the principle in question is sound. Most obviously, we need a positive case for signing up to any harm principle at all. That is a task for another day. The task here is to get clear about which harm principles one might sign up to, what is and is not compatible with signing up, and what objections one is really exposed to if one chooses to do so.

I.

The first four sections of this paper defend the following thesis:

> *The Disambiguation Thesis* (DT): though many write of "the harm principle," there is no single such principle; there are several harm principles.

DT does not claim that any harm principle is sound. It is agnostic on this issue. DT merely claims that though it is common to see discussions of "the harm principle," there are many distinct harm principles in the philosophical literature.[1] To see that this is so, consider two different forms a principle may take:

1) x is only permissible if y;
2) x is not only permissible if y.

A principle taking form 1) is a *constraint*: if y does not obtain, x is not permissible.[2] A principle taking form 2) is *permissive*: if y does not obtain, x may still be permissible.[3] While 1) identifies a necessary condition of the permissibility of x, it says nothing about whether that condition is also sufficient. So we can further distinguish between 1a) and 1b):

1a) y is a necessary but insufficient condition of the permissibility of x;
1b) y is a necessary and sufficient condition of the permissibility of x.

Could a plausible harm principle take form 1b)? It could not. There are further conditions that must be met before any form of coercive regulation

[1] I do not claim that the existence of distinct harm principles has gone unnoticed. But if I am correct, there are many more harm principles than is typically acknowledged.
[2] I take for granted that an act is permissible if it is not wrong or if it is wrong but justified.
[3] Permissive principles, then, negate candidate constraints.

https://doi.org/10.1017/S135232521500004X Published online by Cambridge University Press

is permissible, conditions that are independent of any harm principle.[4] Permissible coercive regulation must also comply with certain demands of the rule of law (publicity, clarity, prospectivity, etc.), with certain basic rights and liberties (privacy, freedom of expression, nondiscrimination, etc.)[5] , and with the requirements of political legitimacy (which may or may not include requirements such as state neutrality,[6] or acceptability to reasonable people)[7] . In what follows, I therefore focus on 1a).

We can also distinguish between 2a) and 2b):

2a) y is a sufficient but not a necessary condition of the permissibility of x;
2b) y is neither a sufficient nor a necessary condition of the permissibility of x.

For reasons given above, no plausible harm principle would take form 2a). What about 2b)? A principle taking this form may seem uninteresting. But if y is widely believed to be a condition of a certain type—if, say, harm prevention is widely believed to be a necessary condition of permissible coercive regulation—a principle that denies this would be well worth exploring. One might still doubt that such a principle would be appropriately described as a harm principle. Some deny that it is a necessary or sufficient condition of permissible coercive regulation that the regulation prevents harm. It does not seem that they thereby endorse any harm principle. However, 2b) is compatible with 2c):

2c) y is neither a necessary nor a sufficient condition of the permissibility of x but is a necessary part of a sufficient condition of the permissibility of x.

Principles that take form 2c) remain permissive: x is not only permissible if y. But they do not merely deny that y is a condition of a certain type: they also make an affirmative claim about the relevance of y to the permissibility of x. Joel Feinberg endorses a harm principle of this kind. In his view, the harm principle identifies an INUS condition: an insufficient but necessary part of an unnecessary but sufficient set.[8] According to Feinberg, there is a presumption in favor of liberty that must be rebutted if coercive regulation is to be permissible and which can be rebutted only when a so-called liberty-limiting principle is satisfied.[9] Feinberg's harm principle

[4] The reference here to coercive regulation is not meant to foreclose the question of which types of regulation (or types of action) are governed by "the harm principle." Coercive regulation is, for now, a placeholder for various possibilities. I discuss some of these below.

[5] True, these rights are usually thought to be qualified—they can, that is, be infringed without being violated. My point is only that violations of these rights are impermissible.

[6] *Compare* Gerald Gaus, *The Moral Foundations of Liberal Neutrality*, *in* CONTEMPORARY DEBATES IN POLITICAL PHILOSOPHY (Thomas Christiano & John Christman eds., 2009), *with* Richard Arneson, *Liberal Neutrality on the Good: An Autopsy*, *in* PERFECTIONISM AND NEUTRALITY (George Klosko & Steven Wall eds., 2003).

[7] *Compare* DAVID ESTLUND, DEMOCRATIC AUTHORITY (2008), at 21–64, *with* Joseph Raz, *Disagreement in Politics*, 43 Am. J. Juris. 25 (1998).

[8] Because those writing on the topic typically refer to "the harm principle," I occasionally adopt their way of speaking. It should always be remembered, however, that there is no single such principle.

[9] JOEL FEINBERG, HARM TO OTHERS (1984), at 9.

https://doi.org/10.1017/S135232521500004X Published online by Cambridge University Press

is one such principle. It states that "the need to prevent harm (private or public) to parties other than the actor is always an appropriate reason for legal coercion."[10] That reason is *insufficient* to render coercive regulation permissible—as mentioned above, there are other conditions that must also be met. And there is at least one *other* sound liberty-limiting principle—what Feinberg calls the offense principle.[11] So the harm principle states neither a necessary nor a sufficient condition of permissible coercive regulation.[12] For Feinberg there are sets of conditions satisfaction of which suffices to make coercive regulation permissible but which do not include the harm principle (they include the offense principle instead). Yet the harm principle *is* a necessary part of other sets of conditions, satisfaction of which also suffices to make coercive regulation permissible. It follows that Feinberg's principle is a permissive principle that takes form 2c) and that it states an INUS condition.

The distinction between 1a) and 2c) already gives us two different harm principles. We just saw that Feinberg's principle takes form 2c). Mill's harm principle, on the other hand, takes form 1a). According to Mill's principle, "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."[13] This is a *constraint*: it is a necessary (but insufficient) condition of the permissible use of coercive power against *p*, that the power is exercised in order to prevent others being harmed.

Another distinction between principles is worth making at this point:

3) y is a reason to x;
4) y merely eliminates a reason not to x.

Principles that hold that 3) is true I call *positive*. Principles that hold that 4) is true I call *negative*. The term *eliminates* is meant to cover two types of case: first, cases where a reason is *cancelled*, such that a fact that had counted in favor of φing no longer does so; second, cases where a reason is *excluded*, such that a fact that counts in favor of φing, and on which an actor had been able to permissibly act, is no longer a fact for which that actor can permissibly act.[14] The distinction between 3) and 4) is put to work in the literature on punishment. So-called negative retributivists think that if I am guilty of culpable wrongdoing, my guilt cancels reasons not to punish me, but is not itself any reason to punish me.[15] So-called positive retributivists think that my guilt is itself such a reason.

[10] *Id.* at 11.
[11] The offense principle states that "it is always a good reason in support of a proposed criminal prohibition that it is probably necessary to prevent serious offence to persons other than the actor." *Id.* at 26.
[12] Feinberg acknowledges this point; *id.* at 10.
[13] JOHN STUART MILL, ON LIBERTY (Elizabeth Rapaport ed., 1978), at 9.
[14] On exclusionary reasons, *see* JOSEPH RAZ, PRACTICAL REASON AND NORMS (1999), at 35ff.
[15] The only reason to punish, for some who hold this view, is the prevention of further wrongdoing. For an overview of the literature, *see* Antony Duff, *Legal Punishment*, The Stanford Encyclopedia of Philosophy (Edward N. Zalta ed., 2013).

TABLE 1

	Y is a reason to x (3)	Y merely eliminates a reason not to x (4)
Y is a necessary condition of the permissibility of x (1a)	Positive constraint	Negative constraint
Y is not a necessary condition of the permissibility of x (2c)	Positive permissive	Negative permissive

By adding the distinction between 3) and 4) to that between 1a) and 2c), we can identify four different harm principles. All four tell us something about the relevance of y to the permissibility of x. The distinction between 1a) and 2c) allows us to distinguish principles by reference to the *logical* role assigned to y. To repeat, a principle is a *constraint* if y is a necessary condition of the permissibility of x; a principle is *permissive* if y is not a necessary condition of the permissibility of x.[16] The distinction between 3) and 4) allows us to distinguish principles by reference to the *normative* role assigned to y. A principle is *positive* if y is a reason to x; a principle is *negative* if y merely eliminates a reason not to x. As Table 1 shows, these distinctions cut across one another. A principle is *positive* and a *constraint* if a) y provides a reason to x, and b) y is a necessary condition of the permissibility of x. A principle is *negative* and a *constraint* if a) y merely eliminates a reason not to x, and b) y is a necessary condition of the permissibility of x. A principle is *positive* and also *permissive* if a) y provides a reason to x, and b) y is not a necessary condition of the permissibility of x. A principle is *negative* and also *permissive* if a) y merely eliminates a reason not to x, and b) y is not a necessary condition of the permissibility of x.

We saw above that Feinberg's harm principle is both *positive* and *permissive*: the prevention of harm is a reason *for* coercive regulation, but coercive regulation can be permissible in the absence of that reason. Andreas von Hirsch and Andrew Simester endorse a harm principle that is a *positive constraint*. Like Feinberg, they think harm prevention supplies "a positive (though not conclusive) ground to prohibit."[17] Unlike Feinberg, they think that prohibitions that fail to prevent harm are impermissible.[18] Mill's harm principle, on the other hand, is most plausibly seen as a *negative constraint*. The fact that my purpose in prohibiting φing is to prevent harm cannot *itself* be thought to generate a reason to prohibit φing. One cannot bring reasons

[16] It is true, of course, that 1a) and 2c) have the additional properties identified above. I suppress these here for ease of exposition.

[17] ANDREW SIMESTER & ANDREAS VON HIRSCH, CRIMES, HARMS, AND WRONGS (2011), at 35.

[18] This may seem inconsistent with the authors' claim that we should accept an independent offense principle. In fact it is not. While they think there is reason to criminalize offensive wrongdoing, Simester & von Hirsch also think this reason insufficient unless the harm principle is satisfied; *id.* at 117–118.

https://doi.org/10.1017/S135232521500004X Published online by Cambridge University Press

into existence so easily. It is more plausible to think that a harm-preventive purpose eliminates objections that would otherwise apply (say, if my purpose were to benefit my political allies or eliminate harmless immoralities). This makes Mill's constraint negative.

II.

The previous section gives us four harm principles. But different harm principles also unpack x and y in different ways. Start with x: What is it that any given harm principle is concerned with the permissibility of? One possibility is that x is any form of coercion, or at least any form of *societal* coercion. This appears to have been Mill's view.[19] Another possibility is that x is any form of legal regulation. One might combine elements of both these views and say that x is any form of coercive legal regulation. Or one might say that x is restricted to criminalization.[20] There are other possibilities. Which should be endorsed depends on the positive case for endorsing any harm principle at all. To see this, consider two arguments: *the grievance argument* and *the clumsiness argument*.[21] The grievance argument relies on a thought about state legitimacy, where a state's legitimacy, let us say, is its "right to impose new duties on subjects by initiating legally binding directives, to have those directives obeyed, and to coerce noncompliers."[22] The thought on which the argument relies is that the state has no such right unless breach of a directive would give some other person a grievance.[23] To this first premise may then be added a second, to the effect that no other person has a grievance against p unless p's conduct harms (or risks harm to) that person.[24] On this view the state has no right to impose legal duties on p, have those duties obeyed, and back them up with coercion merely

[19] Mill writes that his harm principle is "entitled to govern absolutely the dealings of society with the individual *in the way of compulsion and control* whether the means used be physical force in the form of legal penalties, or the moral coercion of public opinion" (emphasis added). *See* MILL, *supra* note 13, at 9.

[20] Officially this is Feinberg's position: *see* FEINBERG, *supra* note 9, at 3. Note that one might further distinguish between actions that are criminal offenses and actions that are crimes, where an action is a crime only if it is a criminal offense and there is no defense available to those who commit it. A given harm principle might be concerned either with the permissibility of the creation of criminal offenses or with the creation of crimes, or with both. I am grateful to Doug Husak and Patrick Tomlin for pointing out this further distinction.

[21] As my aim here is not to evaluate the arguments themselves, I offer but a sketch of each.

[22] John Simmons, *Justification and Legitimacy*, 109 Ethics 739 (1999), at 752.

[23] One might endorse this first premise for a number of reasons. To take just one, perhaps the state has the right to φ only if its people have consented to its having that right or would consent in certain hypothetical circumstances. And perhaps people have consented or would consent to the state having the right described in the text only in cases in which their conduct aggrieves another.

[24] The argument sketched in the text supports a harm principle that is a constraint. If one thinks that the list of grievances is longer—perhaps a person also has a grievance against p if p's conduct offends that person—one will endorse a version of the grievance argument that supports a permissive harm principle.

because breach of those duties would be immoral. Unless *p*'s breach would cause (or risk) harm to someone else, this is impermissible.

I do not here offer any defense of the grievance argument.[25] But if this is the argument for endorsing some harm principle, it has implications for the identity of x. Most obviously, the argument says nothing about the actions of private individuals. It also says nothing about extralegal state action or about legal regulation that is not coercive. What it does say is that coercive legal regulation of φing is impermissible if φing is not harmful to others. The argument thus supports a harm principle in which x is coercive legal regulation.[26]

Consider now the clumsiness argument. This argument relies on a more mundane thought about the nature of the tools we have available for the regulation of conduct. More precisely, it relies on the effects—specifically the harmful effects—of some of those tools.[27] Take the criminal law as an example. Perhaps the most obvious of the harms done by the criminal law are the harms done by criminal punishment.[28] These *punitive harms* are of different types. Some are *internal* to the punitive exercise—they are harms that must be imposed if the punishment is to succeed qua punishment. Thus a life sentence that does not worsen the prisoner's prospects in life is plausibly thought of as a punitive failure. Perhaps not all punishments are like this. Some fixed penalties may be designed to deprive the offender of resources whether or not this causes any harm. In such cases, harming the offender is not the object of the exercise.[29] But the fixed penalty may nonetheless be harmful to those on low incomes—it may deprive them of scarce resources without which their prospects in life are diminished.[30] Because such harms are an unintended side effect of the fixed penalty, they are one type of *external* punitive harm.

[25] Gerald Postema argues that the grievance argument, or something like it, is implicit in Joel Feinberg's work, particularly Joel Feinberg, Harmless Wrongdoing (1988), at 321–322. *See* Gerald Postema, *Politics Is about the Grievance: Feinberg on the Legal Enforcement of Morals*, 11 *Legal Theory* 293 (2005).

[26] One might, of course, endorse other versions of the grievance argument. One might, for instance, argue that *all* coercion is permissible only if the person coerced has aggrieved, or is about to aggrieve, another. If this argument is sound, there will be a sound harm principle in which x is any coercive act.

[27] It is thus necessary to spend some time discussing these effects. The clumsiness argument itself is set out in the final paragraph of this section.

[28] While it is not inevitable that if φing is criminalized, people will be punished for φing, cases where no one is punished are not typical. I thus follow Douglas Husak in assuming that one cannot ignore the morality of punishment when assessing the morality of criminalization; *see* Douglas Husak, Overcriminalization (2008).

[29] Can one punish without intending to harm? Some argue that one cannot; *see, e.g.*, David Boonin, The Problem of Punishment (2008), at 6–17. For present purposes, this would show only that not all criminal sentences are punishments. It would not alter the main point in the text, namely, that criminalization results in a variety of different types of harm.

[30] Here I assume that one is harmed by that which diminishes one's prospects in life. Those who endorse different conceptions of harm can alter the example as required. The point—namely, that criminal sentences often harm as a side effect—remains.

Internal and external punitive harms are by no means mutually exclusive. Even when a punishment is internally harmful, it may impose substantial external harms. Some such harms occur during punishment—those serving prison sentences may be abused by fellow prisoners, suffer psychological trauma, or develop debilitating addictions. Other such harms occur only later—the punished may be unable to resume careers or reestablish family relationships. Nor are external harms suffered only by the punished. The lives of families, friends, and employers can also be blighted by punishment, most obviously when relationships of all these kinds are damaged or destroyed.

The aforementioned punitive harms comprise one subclass of the *ex post* harms brought about by criminalization. These harms are *ex post* in the sense that they occur downstream of the commission (or suspected commission) of a crime. Other such harms are easily forgotten. Some are the product of aborted investigations or of trials that end without convictions: if the accusation of criminality sticks, *p*'s prospects may be changed for the worse. Others are the product of convictions unaccompanied by punishments: a criminal record may damage *p*'s employment prospects even if he is discharged unconditionally.

In addition to these *ex post* harms, criminalization may also be harmful *ex ante*—prior, that is, to the suspected commission of any crime. Much of this *ex ante* harm is a product of the message sent by criminal prohibition, namely, that the prohibited conduct is not to be done and that those proven to have offended will be punished.[31] If φing is criminalized, this message is likely to lead at least some people to avoid φing. To give themselves assurance against being prosecuted, some may also avoid μing—conduct in some way connected to φing that they fear may also lead to prosecution. One may wonder in what sense these reactions could be harmful. But as already suggested, on a plausible view one is harmed by that which diminishes one's prospects in life.[32] One's prospects in life are partly a function of one's ability to pursue whatever valuable projects and goals one has set for oneself. If μing is a constituent of one of *p*'s projects, and *p* no longer feels able to pursue that project because of a new criminal offense, *p*'s prospects are adversely affected by criminalization.[33]

The same goes for cases where φing is the act-type in question. One might object that if lawmakers act permissibly, only wrongful acts will be

[31] This is the case unless, of course, they can prove a defense. That this message is sent by criminal offenses is seen by Hart. *See* H.L.A. HART, THE CONCEPT OF LAW (3d ed. 2013), at 28.

[32] Some think that *p* is only harmed by Z if *p*'s prospects in life are caused to be *bad* by Z, such that a very well-off person whose prospects are diminished may not have been harmed if those prospects remain fairly good. Again, my examples could easily be altered without cost to the argument.

[33] Does it matter whether *p*'s reaction to criminalization is reasonable? Is *p*, who unreasonably refuses to φ, harmed by events that cause her unreasonably to refuse to φ? Perhaps not. But the types of reaction described in the text need not be unreasonable, so we need not resolve this point here.

criminalized, and that if φing is wrongful, it cannot be a constituent of a valuable project or goal; the fact that people feel unable to φ cannot then detract from their prospects. This line of thought harbors several errors, two of which are worth mentioning here.

First, the line of thought is most plausible in relation to *mala in se*—wrongs that are wrongs prior to or independent of the law. But when lawmakers target such wrongs, they will often (perhaps always?) create offense definitions that are overinclusive. Why? Because offense definitions are rules, and it is (all but?) impossible to avoid creating rules that are over- and underinclusive relative to targets that preexist them.[34] Prohibited acts that were *not* wrongful prior to their regulation may well be constituents of valuable projects and goals.

Second, and more important, it does not follow from the fact that φing is wrongful that φing cannot be part of a valuable project or of a valuable life more generally. Take a religious group that practices forcible circumcision of the children of group members at the age of ten. Let us assume that those who do not consent are wronged if they are circumcised. This group's way of life may nonetheless have many valuable aspects, and members of the group (and others) may believe the wrong to be integral to it. In such a case, there is a standing risk that criminalizing the wrongful practice will be seen (by members of the group and others) as expressing more general condemnation of the group's way of life.[35] Such condemnation may have various detrimental effects, not least on the self-respect of group members. As John Rawls points out, such self-respect is not to be underestimated: "Without it nothing may seem worth doing, or if some things have value for us, we will lack the will to strive for them."[36] Criminalizing a wrong may thus do much *ex ante* harm.

No doubt this is only some of the harm brought about by criminalization. It is true that some criminal offenses may not bring about all the types of harm I mention. But any offense is highly likely to bring about at least some of them. Those who create offenses also have relatively little control over which harms each offense will bring about. True, they can control whether φing is prohibited rather than μing, and what range of punishments is available to the courts. But who ends up being harmed depends on the decisions of law-applying officials (who will be prosecuted and convicted? who will be punished and how much?) and private individuals (how

[34] Sometimes the targets of a rule are themselves constituted by the rule. I set such cases aside in the text. For discussion of the nature of rules, *see* FREDERICK SCHAUER, PLAYING BY THE RULES (1991); LARRY ALEXANDER & EMILY SHERWIN, THE RULE OF RULES (2001).

[35] I draw here on the work of Joseph Raz; *see* JOSEPH RAZ, *Free Expression*, *in* ETHICS AND THE PUBLIC DOMAIN (rev. ed. 1994), at 160–162. As Raz points out, it is no response to say that one's *intention* is not to condemn the group's way of life or that the wrongful practice is detachable from the valuable parts of that way of life. We have limited control over the social significance of our acts, and it is presumably for individuals, not the state, to decide what is integral to a particular way of life.

[36] JOHN RAWLS, A THEORY OF JUSTICE (rev. ed. 1999), at 386.

will potential offenders react to the new offense, and how will those they know react to their prosecution or conviction or punishment?). Nor can any official ensure that whatever harms ensue are suffered only by the deserving.[37] Because mistakes are inevitable in any system of criminal justice, not all punishments will be imposed on those who φ. Because overinclusion is (all but?) inevitable in any system of rules, not all φers will deserve to be harmed. And because the just treatment of offenders typically requires that sentencers be bound by rules,[38] there will be overlong sentences as well as overinclusive offenses, so even those who deserve to be harmed will not get precisely what they deserve. Not all internal punitive harms will thus be deserved in any actual legal system. And there are external punitive harms, other *ex post* harms, and *ex ante* harms too.

This is enough to show that criminalization is a tool that is clumsy in a particular sense: those who put it to use cannot control the harm it will do in the course of achieving their objectives (if it does), but they do make it highly likely that some harm will be done, and all but inevitable that when harm is done, some of it will be undeserved. The crucial premise, of course, remains to be stated: it is that putting such a tool to use requires a particular type of payoff—a payoff in terms of harms prevented rather than other good things.[39] I cannot defend that premise here. The main point is rather that the argument again has implications for x. If the clumsiness argument is the argument for some harm principle, it applies only to tools that are clumsy in the sense just described. In principle, these may or may not be legal tools and they may or may not be coercive. When it comes to x, the implications of the clumsiness argument are thus very different from that of the grievance argument, producing yet more different harm principles.

[37] Or those otherwise morally liable to suffer harm. I focus on the narrower category of desert in the text, but the point seems to me to hold on any plausible view of liability, where *p* is morally liable to be harmed to some degree if *p* lacks a moral right not to suffer that harm.

[38] If only to ensure a degree of consistency between offenders who are similarly placed.

[39] One might reply that in a just legal system, punitive harms are deserved by those on whom they are inflicted and that inflicting deserved harm is an unqualified good. It might be said that the state has reasons of justice to impose said harms, not reasons to avoid imposing them for which it must struggle to compensate; perhaps conforming to the former reasons can sometimes justify criminalization even if no harm principle is satisfied. This can be doubted. Even if there are impersonal reasons to impose harm on offenders who deserve it—reasons that derive from the value of doing retributive justice—there may also be personal reasons not to impose harm on those same offenders—reasons that derive from the fact that all harm is bad for the person harmed. If this is right, even deserved punitive harms are harms the state has reason to avoid. And this makes it much less likely that our reasons to impose deserved harms could themselves defeat the reasons not to criminalize given by the other harms that criminalization brings about. For discussion of the relationship between retributivist theories of punishment and various harm principles, *see* Patrick Tomlin, *Retributivists! The Harm Principle Is Not for You*, 124 Ethics 272 (2014).

https://doi.org/10.1017/S135232521500004X Published online by Cambridge University Press

III.

Section I identifies four harm principles. Each can be stated in the form "y is a condition of the permissibility of x." Having considered some candidates for x, we now turn our attention to y. To simplify, I assume here that x is criminalization.[40] What, then, is it that different harm principles claim is a condition of permissible criminalization? Consider three possibilities:[41]

5) criminalization of φing will probably prevent harm;
6) φing is harmful;
7) φing is criminalized in order to prevent harm or to prohibit what is harmful.

5) interprets y *instrumentally*. It is the likelihood of actual harm prevention that is a condition of permissible criminalization. It is plausible to think that a complete statement of 5) would add conditions relating to alternative means of prevention and to considerations of proportionality. I return to this topic in Section X. Leaving such additional conditions aside, 5) might be amended in at least two ways. Consider:

5a) criminalization of φing will prevent harm;
5b) criminalization of φing will (probably) prevent harm being caused by those who φ.[42]

Both 5a) and 5b) also give y an instrumental interpretation. But each identifies a condition that is more difficult to satisfy than 5). If criminalizing φing is likely to prevent harm being caused by those who commit no offense or by natural causes, this is enough to satisfy 5). But it is not enough to satisfy 5a) or 5b).[43] Sections VII to XI argue that a harm principle in which 5) figures as a condition of permissible criminalization is a principle that survives a number of objections. Whether 5a) and 5b) also survive is a question for another day.

6) gives y an *act-centered* interpretation. It is the fact that the conduct criminalized is harmful that is a condition of permissible criminalization.[44] 6) might be amended to give other act-centered conditions. Consider:

[40] For present purposes, φing is criminalized only if φing is a criminal offense, and φing is a criminal offense only if those proved to have φed are liable to be convicted and sentenced in a criminal court unless they are able to prove a defense.

[41] I do not claim these are the only three, but they are prominent in the literature and worth considering for that reason.

[42] The distinction between 5) and 5b) is noted by others. *See* Nils Holtug, *The Harm Principle*, 5 Ethical Theory & Moral Prac. 357 (2002), at 360; David Brink, Mill's Progressive Principles (2013), at 184.

[43] Whichever version of 5) we accept, satisfaction of the condition in question can be blocked by other criminal offenses that already exist in the jurisdiction. If preexisting offenses prevent harms that criminalizing φing might otherwise have prevented, criminalizing φing will not be likely to prevent harm, and 5) will not be satisfied. Of course, if we endorse a harm principle that is a constraint, those preexisting offenses must also satisfy 5). If they do not, they should be repealed.

[44] One question for any defender of 6) concerns the relationship that must exist between a given action and a given harm for that action to count as harmful. Must my action have caused the harm, and in what sense of "caused"? For discussion, see Andrew von Hirsch, *Extending the*

https://doi.org/10.1017/S135232521500004X Published online by Cambridge University Press

6a) φing creates a risk of harm the creation of which is unjustified.

Because conduct that creates an unjustified risk of harm is not always harmful, 6a) identifies a condition that is satisfied in some cases in which 6) is not. Because conduct that is harmful does not always create an unjustified risk of harm, 6) identifies a condition that is satisfied in some cases in which 6a) is not. I return briefly to 6a) in section VII. 7) interprets y *purposively*. It is the fact that the creator of the criminal offense in question *aims* to prevent harm, or prohibit what is harmful, which is a condition of permissible criminalization.

The harm principles given by 5), 6) and 7) are importantly distinct. Take the act-centered and instrumental principles first. There is no inevitable connection between criminalizing an act-type and reducing its incidence or the harm done by it. It is entirely possible that the main effect of criminalizing φing is that φing becomes more appealing to people who would otherwise have eschewed it and is driven into the hands of unscrupulous souls who make it even more harmful. So even if φing is harmful, there is no guarantee that criminalizing φing will—or is likely to—prevent harm. The condition in 6) may thus be satisfied even where the condition in 5) is not.

The reverse is also true. There may be harms that can be prevented only by criminalizing harmless acts. Take the offense of battery, which in English law criminalizes all unconsented-to physical contact not in keeping with expectations appropriate to the time and place.[45] Many tokens of this act-type are not harmful.[46] Criminalizing all batteries thus does not satisfy the condition in 6). But it may satisfy 5). It will do so if, were all batteries *not* criminalized, there would likely be more harmful batteries.[47] Whether this would be so is, of course, an empirical question. But if we decriminalized harmless batteries and this were widely publicized, it is plausible to think that more people would act in ways they believed would amount to harmless battery. It is plausible to think that some of these additional batteries would in fact cause harm (misjudgment and escalation are only two possibilities). This is harm that criminalizing all batteries would have prevented.

The previous paragraph assumes that 6) requires that each token of a prohibited act-type is harmful. Other versions of 6) are possible. Perhaps those tokens must tend to be harmful (e.g., dangerous driving) or must be harmful when considered together (e.g., pollution).[48] Note, however,

Harm Principle: Remote Harms and Fair Imputation, *in* Harm and Culpability (Andrew Simester & A.T.H. Smith eds., 1996).

[45] Collins v. Wilcock, [1984] 3 All E.R. 374.

[46] As I discuss below, there are various senses of the term *harm*, but I take for granted here that not all unwanted and unusual touching causes any harm. Nor does it create an unjustified risk of harm.

[47] Again, this is subject to any additional conditions built into a full statement of 5) but which I temporarily suppress above.

[48] These possibilities are noted by Gardner & Shute; *see* John Gardner & Stephen Shute, *The Wrongness of Rape, in* Oxford Essays in Jurisprudence: Fourth Series (Jeremy Horder ed., 2000).

TABLE 2

	Constraint (1a)	Permissive (2c)
Instrumental (5)	HP1	HP2
Act-centered (6)	HP3	HP4
Purposive (7)	HP5	HP6

that each of these versions of 6) remains act-centered. If we endorse such principles, what matters to permissibility are the consequences of the acts we are considering criminalizing. If we endorse an instrumental principle, what matters to permissibility are the consequences *of criminalization itself*. If we endorse a constraint in which y is 5) and discover that criminalization is an entirely ineffective means of preventing harm, our chosen principle renders it impermissible to criminalize anything; if y is 6), this discovery is irrelevant to what our chosen principle permits.[49]

So act-centered and instrumental principles cover different ground. And both are distinct from purposive principles.[50] That my aim is to prevent harm does not mean I am actually likely to prevent any; and a criminal offense may prevent harm whether or not the aim of its maker was to do anything of the sort. Similarly, that my aim is to prevent harm does not mean the conduct I prohibit is necessarily harmful; and a criminal law may prohibit conduct that is harmful whether or not its maker's aim was to prohibit harmful acts.

Suppressing temporarily the distinction between positive and negative principles, and holding x constant, we can now identify the harm principles set out in Table 2.

I already pointed out that Mill's harm principle is both purposive and a constraint. Mill thus endorses HP5. As mentioned above, Simester and von Hirsch's harm principle is a constraint. It is also instrumental: the proponents of criminalization must "supply evidence that creating a new offence *will* help to prevent harm."[51] This is to endorse HP1. That said, Simester and von Hirsch are less careful than they might be about the distinction between instrumental and act-centered principles. They also write that "the point of the Harm Principle is that it forces an enquiry into the consequences of conduct" and requires that we ask "does it hurt

[49] The difference between 5) and 6) is noted by others; *see* Victor Tadros, *Harm, Sovereignty, and Prohibition*, 17 Legal Theory 35 (2011); Gardner & Shute, *supra* note 48; Tomlin, *supra* note 39.

[50] For some worries about purposive harm principles, *see* Holtug, *supra* note 42, at 362–363.

[51] SIMESTER & VON HIRSCH, *supra* note 17, at 35, 47–48 (emphasis in original). HP1 is also endorsed in Gardner & Shute, *supra* note 48.

https://doi.org/10.1017/S135232521500004X Published online by Cambridge University Press

anyone"?[52] This is to switch to an act-centered principle, HP3. As just shown, HP1 and HP3 are not the same principle.[53]

We already saw that Joel Feinberg endorses a permissive principle. In its canonical formulation, the principle is instrumental: "it is always a good reason in support of penal legislation that it will probably be effective in preventing (eliminating, reducing) harm to persons other than the actor."[54] This is to endorse HP2. Douglas Husak, on the other hand, endorses HP6. Husak writes that criminal liability "may not be imposed unless statutes are designed to prohibit a nontrivial harm *or evil*."[55] He calls this the nontrivial harm or evil constraint. The reference to what statutes are "designed" to achieve suggests that Husak's principle is purposive—what matters are the aims of those who create a given offense. Viewed as a harm principle, Husak's principle is also permissive: criminalization may be permissible even if the lawmaker's aim is not to prohibit the harmful but to prohibit evil yet harmless acts.

Note that the positive case for endorsing some harm principle has implications for y as well as x. According to the grievance argument, the state has a right to criminalize φing only if φing generates a grievance, and φing generates a grievance only if φing is harmful (or perhaps offensive) to others. The argument thus supports HP3 (or perhaps HP4). According to the clumsiness argument, the clumsiness of tools like the criminal law gives us reasons not to use them, reasons that can be defeated only if their use will probably prevent harm. This argument thus supports HP1.

IV.

The preceding sections are enough to vindicate the disambiguation thesis, DT. In the following sections, I explore some payoffs of the disambiguation for the soundness of different harm principles. Suffice it to say here, however, that further distinctions could be drawn. I mention just two.

[52] SIMESTER & VON HIRSCH, *supra* note 17, at 35.

[53] Simester & von Hirsch's discussion of so-called remote harms may also seem to suggest that they endorse an act-centered principle. Remote harm occurs when D's action leads to harm only via some "contingency," such as a "mediating intervention" by some other person. The authors claim that cases of this type "do not fit neatly within the Harm Principle," such that an "extended" version of that principle is required; such a principle permits criminalization of remotely harmful acts, but only where D can be held responsible for the harm; *id.* at 53–88. Note, however, that this worry about the type of connection needed between D's actions and particular harms is a worry only if one endorses an act-centered principle. If one endorses an instrumental principle, the relevant question is not whether the conduct in question brings about harm but whether criminalization of that conduct is likely to prevent harm. Because Simester & von Hirsch explicitly sign up to an instrumental principle, their discussion of remote harms is best interpreted as of relevance not to the harm principle but to another principle they endorse, the so-called *wrongfulness constraint*. I discuss this constraint in Section V.

[54] FEINBERG, HARM, *supra* note 9, at 26.

[55] HUSAK, *supra* note 28, at 66 (emphasis added).

https://doi.org/10.1017/S135232521500004X Published online by Cambridge University Press

The first distinction is given by our answer to the following question: Must the harm to which any harm principle refers be harm done *by* one or more persons *to* one or more others?[56] If we answer "yes," we endorse a *narrow* harm principle. According to a narrow version of HP3, φing is permissibly criminalized only if φing is harmful to others. If we answer "no," we endorse a *wide* harm principle.[57] According to a wide version of HP3, φing is permissibly criminalized only if φing is harmful to a person. This principle thus does not exclude criminalization in cases where φing is harmful only to the very person who φs.[58]

A second distinction is given by different conceptions of what it is to be harmed. There are many such conceptions, and I can only scratch the surface here. Some are *comparative*.[59] On these views, *p* is harmed by Z if *p* is made worse off by Z in some respect and for some period of time, relative to some baseline. Comparative conceptions must thus explain both a) in what respect *p* must be made worse off if *p* is to be harmed, and b) relative to what baseline. Take b) first. If we adopt a *temporal* baseline, the relevant baseline is the position *p* was in the moment before Z occurred. On this view, if Dr. A maliciously prolongs *p*'s fever by administering some drug, Dr. A does not harm *p* thereby: *p* is no worse off than she was the moment before administration. If we adopt a *counterfactual* baseline, the relevant baseline is the position *p* would have been in but for Z. On this view, everything turns on what would have happened if Dr. A had not administered the drug. If Dr. B would have poisoned *p* had Dr. A not acted, *p*'s baseline position is that in which she has been poisoned. Assuming it is better (or no worse) to have the fever than be poisoned, *p* is not harmed by Dr. A.[60] If we instead

[56] I leave open whether these must be other persons or whether the others in question can also be nonhuman animals.

[57] Joseph Raz is one writer who endorses a wide harm principle. *See* Joseph Raz, *Autonomy, Toleration and the Harm Principle, in* Issues in Contemporary Legal Philosophy: The Influence of H.L.A. Hart (Ruth Gavison ed., 1987), at 326.

[58] Some will object that only narrow principles warrant the label "harm principle." They may point to Mill's harm principle and to the tradition of thought that considers the harm principle a distinctively liberal principle. In brief reply, the discussion already shows that Mill's is only one harm principle among several others. Both wide and narrow principles make harm a direct determinant of the permissibility of criminalization (or legal regulation or coercion); there is thus nothing obvious about denying them the label "harm principle." Nor are the liberal credentials of a principle obviously imperiled if it is wide. HP3 still rules out criminalization of harmless wrongs, and HP1 still requires that criminalization prevent harm. Such harm principles thus remain incompatible with many legal moralist principles, an incompatibility that gives the harm principle much of its liberalizing appeal.

[59] I focus on these conceptions because I think them most plausible, though I cannot defend this here. For others, *see* Seana Valentine Shiffrin, *Harm and Its Moral Significance*, 18 Legal Theory 357 (2012); Matthew Hanser, *The Metaphysics of Harm*, 77 Phil. & Phenomenological Res. 421 (2008). For criticism of Shiffrin, *see* Victor Tadros, *What Might Have Been, in* Philosophical Foundations of the Law of Torts (John Oberdiek ed., 2014). For criticism of Hanser, *see* Judith Jarvis Thomson, *More on the Metaphysics of Harm,* 82 Phil. & Phenomenological Res. 436 (2011).

[60] Judith Jarvis Thomson offers an improved version of this view. On Thomson's view, the relevant baseline is the position that Z prevents *p* from being in, by the very means by which Z affects *p's* position. The means by which Dr. A affects *p* is administration of the drug. This means does not itself prevent Dr. B poisoning the patient but does itself prevent the patient

adopt a *normative* baseline, the relevant baseline is the position in which *p* is morally entitled to be. The actions of both Doctors A and B violate *p*'s rights. Assuming each action makes *p* worse off in the sense relevant to a), *p* is harmed by whichever occurs.[61]

It is possible to combine these baselines. Thus Raz writes that "one harms another when one's action makes the other person worse off than he was, or is entitled to be."[62] Feinberg, on the other hand, denies that one is harmed by Z (in the sense relevant to the harm principle) if Z does not violate one's rights.[63] I take no position in this debate here. Whatever the answer, comparative accounts must also say something about a). On the widest possible view, *p* is harmed if he is made worse off in any respect, for any period of time. If rights violations make one worse off, and we opt for a normative baseline, then any rights violation is harmful. On a more restrictive view, harm has a prospective dimension: *p* is harmed by *φ*ing only if *φ*ing diminishes *p*'s prospects in life.[64] Typically this occurs when *p* is deprived of valuable opportunities (closure of local factories deprives some of the opportunity to secure employment) or of the ability to use them (if *p*'s arms are broken, *p* cannot work at the factories that remain). On this view, many things that make one worse off are not necessarily harmful:

> a visit to the dentist may be painful but usually it does one no harm. Being frisked at the airport is usually a harmless indignity, and having to wait in a queue is usually a harmless inconvenience. One may be offended by jokes in bad taste, or insulted by rude remarks.[65]

Why need there be no harm here? Because pain, indignity, inconvenience, offense and insult need not deprive one of opportunities, or the ability to use them. True, each may do so. One may be excluded from the life of the workplace by offensive jokes, or rendered incapable of going to the dentist by fear of the pain. Once we identify these further detrimental effects we have identified some harm. This is not to say that insults and offensive jokes are not already *wrongs*. It is merely to say that they are wrongs irrespective of whether they are harmful.

from recovering. So on this view, *p*'s baseline position is one of recovery, and Dr. A does harm the patient. *See* Thomson, *supra* note 59. A complex counterfactual view is also defended in Tadros, *What Might*, *supra* note 59.

[61] For objections to various baselines, *see* Hanser, *supra* note 59. For an argument that whatever baseline one adopts, one ends up with an implausible version of the harm principle, *see* Holtug, *supra* note 42. I cannot address Holtug's argument here.

[62] JOSEPH RAZ, THE MORALITY OF FREEDOM (1986), at 414.

[63] FEINBERG, HARM, *supra* note 9, at 34ff.

[64] *See* Raz, *Autonomy*, *supra* note 57, at 327. *See also* JOHN GARDNER, OFFENCES AND DEFENCES: SELECTED ESSAYS IN THE PHILOSOPHY OF CRIMINAL LAW (2007) at 4, 30, and 244; SIMESTER & VON HIRSCH, *supra* note 17, at 36. This view is criticized in Tadros, *What Might*, *supra* note 59.

[65] John Gardner & James Edwards, *Criminal Law*, *in* The International Encyclopedia of Ethics (Hugh LaFollette ed., 2013).

https://doi.org/10.1017/S135232521500004X Published online by Cambridge University Press

There are further options that lie in between these views.[66] It has been objected, however, that *whichever* view one adopts, harm principles that are constraints can be defended only if the defender adjusts her view of harm in a way that renders those principles "vacuous."[67] The objection runs as follows: there are cases in which it seems that φing *is* permissibly criminalized but where the aforementioned harm principles do not permit criminalization on a plausible conception of harm. If these harm principles are to be saved, that conception must hence be adjusted. But if we make the adjustment whenever we think φing is permissibly criminalized, the harm principle simply tells us that what we think is permissibly criminalized is permissibly criminalized. I argue in Section VII that this argument fails: one can stick to a plausible conception of harm without giving up on harm principles that are constraints.

V.

So much for the first thesis. My second thesis can be defended more quickly.

> *Compatibility Thesis* (CT): endorsement of many harm principles is compatible with endorsement of many other principles that impose conditions on x.

"Other principles" are those that use concepts other than harm (such as wrongfulness, offense, or sovereignty) as the direct determinants of what is (im)permissible. CT claims that we can consistently endorse many of these principles alongside many harm principles. This may seem obvious.[68] But I argue below that the truth of CT helps undermine several prominent objections to the harm principle.[69] It is true, of course, that the compatibility is limited. Take 8):

[66] One such view is Feinberg's, for whom p is harmed only if one of p's interests is adversely affected. A Feinbergian interest can be adversely affected without this diminishing the interest-holder's prospects in life (by, e.g., a trespass to land that goes unnoticed by the owner), and not everything that makes one worse off adversely affects a Feinbergian interest (including that which merely distresses, offends, or irritates). *See* FEINBERG, HARM, *supra* note 9, at 45ff. For other views, *see* Tadros, *What Might*, *supra* note 59.

[67] *See* ANTONY DUFF, ANSWERING FOR CRIME: RESPONSIBILITY AND LIABILITY IN THE CRIMINAL LAW (2008), at 130.

[68] Indeed, it is already implicit in the claim, made in Section I, that any harm principle is only plausibly a necessary but insufficient condition of permissible coercive regulation.

[69] A failure to grasp that truth also seems to be responsible for Feinberg's claim that "any plausible formulation of the harm principle" requires a special sense of harm that employs a normative baseline; *see* FEINBERG, HARM, *supra* note 9, at 34. Feinberg seems to think this because he thinks that no act should be criminalized if it does not prevent violations of some moral entitlement. Even if this is so, it shows only that as well as a harm principle, we should also endorse another principle according to which actions should be criminalized only if they prevent violation of a moral entitlement. It does not show we should invent a special conception of harm.

8) it is permissible to criminalize φing in order to bring about the condemnation of moral wrongdoers in a manner consistent with the rule of law.[70]

If, like Mill, one endorses HP5, one cannot simultaneously endorse 8).[71] Things are different if one endorses HP6, because HP6 is permissive, and thus leaves open that it *may* be permissible to criminalize for reasons like that mentioned in 8).[72] This is enough to show that permissive principles, like HP6 and 8), are compatible with one another. So are many constraints. Consider 9):

9) that φing is morally wrong is a necessary condition of the permissibility of criminalizing φing.

This principle is an act-centered constraint. But it is compatible with HP3, another act-centered constraint. To endorse HP3 and 9) is to claim that if φing is to be permissibly criminalized, it must not only be a moral wrong, it must also be harmful.[73] 9) is also compatible with the instrumental constraint, HP1. To endorse HP1 and 9) is to claim that φing must be a moral wrong and that criminalization of φing must also be likely to prevent harm.

So harm principles that are constraints, like HP1, are compatible with other constraints. We should add that the claims just made about harm principles and other principles are also true of harm principles *and other harm principles*. To take just one example, one can endorse HP1 and claim that criminalization must prevent harm while also endorsing HP5 and claiming that the purpose of those who criminalize must be to prevent harm.

As with DT, the payoffs of CT are explored in due course. Before moving on, it is worth noting one way in which harm may be relevant to the application of principles that are not harm principles. To see this, return to 9). To apply 9), one needs to determine whether any given act-type is a moral wrong. In some cases, one cannot do so without considering the harm that tokens of that type may do. Act-types are wrongs if there is a duty not to commit tokens of the type. In cases where the tokens of a type are necessarily harmful (say, murder), that harmfulness will often be decisive in explaining why there is a duty not to commit them. The same goes for act-types the tokens of which necessarily create a risk of harm (say, dangerous driving). But even in cases where some tokens of an act-type are not themselves harmful nor creative of a risk of harm (the possessor of firearms

[70] As in the previous sections, I assume here that x is criminalization. It is true, of course, that the state could, in principle, condemn wrongdoers for φing without ever prohibiting φing, but it is plausible to think that the rule of law demands the prohibition as a condition of the permissibility of condemning φers.

[71] HP5 states that φing is permissibly criminalized only if φing is criminalized in order to prevent harm or to prohibit what is harmful.

[72] I emphasize the word "may," because while HP6 is permissive, it need not permit lawmakers to criminalize with just any aim in mind. There may be a short list, which may or may not include the aim mentioned in 8).

[73] Of course, if one endorses a normative baseline, that φing is harmful already entails that φing is morally wrong.

https://doi.org/10.1017/S135232521500004X Published online by Cambridge University Press

who locks them in an unbreakable safe, the expert driver who speeds on a deserted open road), the fact that some tokens do create a risk of harm (the possessor of firearms with an itchy trigger finger, the novice driver who speeds down country lanes) may be crucial to the wrongness of those that do not.

The truth of this last claim depends on the following observation: It may be possible to prevent some harms only by prohibiting all tokens of some act-type. If novices think themselves expert drivers, a prohibition of dangerous driving alone is unlikely to change their habits—after all, they think they are driving safely. But if all speeding is prohibited, some may abide by the limit—they may be aware that 50 miles per hour is not 30 miles per hour and may find sufficient reason to comply with the law. If the burdens imposed on those who create no risk are reasonably low, this harm-preventive payoff may suffice to make the prohibition morally binding on those to whom it applies. If the prohibition is morally binding, it is wrong to violate it—the legal duties imposed by the prohibition are moral duties too. Even those whose acts create no risk of harm will thus act wrongly when they violate the prohibition.[74] Because the prohibition is justified in part by its harm-preventive properties, the wrongness in question derives in part from harmfulness.[75]

VI.

In the following sections I consider four objections to "the harm principle." The first I call the *underinclusiveness objection* (UO). According to UO, the harm principle rules out criminalization of actions that—we would all accept—can permissibly be criminalized. Clearly, UO does not apply to permissive principles such as HP2, HP4, and HP6. I argue that it also does not succeed as an objection to at least one harm principle that is a constraint. Some would claim that even if this is correct, harm principles that avoid UO are subject to three further objections. First, those principles misrepresent the reasons why many act-types ought to be criminalized. Call this the *wrong kind of reasons objection* (WO). Second, those principles conflict with the so-called means principle, which states that it is impermissible to use p as a mere means of achieving one's own ends or the ends of others. Call this the *mere means objection* (MO). Third, those principles are overinclusive: they permit the criminalization of actions that—we would all accept—cannot

[74] For similar thoughts on this issue, *see* SIMESTER & VON HIRSCH, *supra* note 17, at 24–29.

[75] There are difficult questions here about the relationship that must exist between a given act and some harm for the latter to contribute to the wrongfulness of the former. It is these questions to which discussions of so-called remote harms are trying to find answers. As Simester & von Hirsch put it, "in the context of remote harms . . . it needs to be shown why the prospect of some ultimate harm, perhaps brought about by an independent actor, makes it wrong for the defendant to act as she does." *See* SIMESTER & VON HIRSCH, *supra* note 17, at 72. If D's act is not wrong, 9) is not satisfied.

permissibly be criminalized. Call this *the overinclusiveness objection* (OO). I argue that the truth of DT and CT helps us see that one harm principle—HP1—escapes both UO and all three further objections.[76]

VII.

A enters B's home without B's permission, takes a nap in B's bed, and leaves. No one learns of what A has done, and A's actions have no discernible effects. Has A wronged B? We can accept that he has. According to Arthur Ripstein, such wrongs are fatal to the harm principle. While Ripstein acknowledges that the harm principle has been interpreted in different ways, he claims that A's wrong exposes a "core difficulty" that "cuts across these differences."[77] The difficulty is that "there are wrongs that any plausible version of liberalism will treat as appropriate objects of criminal sanction, but that are not harmful on any plausible conception of harm."[78] Ripstein claims that because the harm principle rules out criminalization of these harmless wrongs, it must be rejected. This is the underinclusiveness objection, UO.

Ripstein is not alone in pressing UO. According to Antony Duff, cases like Ripstein's present defenders of the harm principle with a dilemma: either they revise their definition of harm, so that B is harmed by A's wrong, or they accept that A's wrong amounts to an exception to the harm principle.[79] The former route renders the harm principle "vacuous"; the latter abandons the harm principle as a constraint.[80]

DT shows us that Duff's is a false choice. We can accept that A's is a harmless wrong. Yet we can also accept a harm principle that is a constraint. We can accept HP1.[81] Are we committed to decriminalizing all harmless wrongs? We are not. As shown above, because HP1 is an instrumental principle, it permits criminalization of harmless acts if criminalizing those actions will probably prevent harm. Both Ripstein's core difficulty and Duff's dilemma assume the harm principle is the act-centered principle HP3: they assume that if there are harmless wrongs that are permissibly criminalized, the harm principle must be rejected, at least as a constraint. This is not the case.[82]

[76] This is not to say that no other harm principle escapes the objections.
[77] Arthur Ripstein, *Beyond the Harm Principle*, 34 *Phil. & Pub. Aff.* 216 (2006), at 217.
[78] *Id.*
[79] DUFF, ANSWERING, *supra* note 67, at 130, 135. For discussion of whether it could be said that A harms B, *see* Ripstein, *supra* note 77, at 218–222; Tadros, *Harm, Sovereignty*, *supra* note 49, at 48. I ignore this avenue of escape here.
[80] DUFF, ANSWERING, *supra* note 67, at 130.
[81] HP1 states that φing is permissibly criminalized only if criminalizing φing would probably prevent some harm. As mentioned, a full statement of HP1 would include additional conditions, on which I say more in Section X.
[82] In a more recent paper, Duff and Marshall acknowledge both the distinction between HP1 and HP3 and that the failure to make that distinction "pervades" the relevant literature. *See* R.A. Duff & S.E. Marshall, *"Remote Harms" and the Two Harm Principles*, *in* Liberal Criminal

https://doi.org/10.1017/S135232521500004X Published online by Cambridge University Press

To test the implications of HP1 for harmless trespass, we must consider two otherwise identical worlds: i) a world in which both harmful and harmless trespass are prohibited, and ii) a world in which only harmful trespass is prohibited.[83] To apply HP1, we must ask whether this difference would likely result in a further difference: there being less harm in world i). It follows that we cannot work out whether HP1 permits criminalization of A's act-token without considering the effect of prohibiting harmless trespass as an act-type.[84] Would such a prohibition satisfy HP1? There is reason to think it would, at least if its existence were widely publicized. Because there would be no criminal liability for harmless trespass in world ii), some people in world ii) would likely trespass when they would not have done so in world i). If some of these trespasses turned out to cause harm (say, because an occupant thought to be on holiday was at home after all), then, ceteris paribus, there would be more harm in world ii) than in world i). What is more, if some people in world ii) felt compelled to expend additional resources securing their homes against trespassers while others who lacked such funds felt less able to enjoy time in their homes, these people might be unable to pursue valuable opportunities they could have pursued in world i). On the prospective view of harm discussed in Section IV, this inability is a type of harm. Pending consideration of alternative means and proportionality, criminalization of harmless trespass would then satisfy HP1.[85]

It is worth noting that defenders of HP3 may also be able to reply to UO. Duff and Ripstein assume that HP3 states that φing is permissibly criminalized only if φing is harmful. But as shown above, there are other versions of that principle. According to one version, φing is permissibly

Theory: Essays for Andreas von Hirsch (A.P. Simester, Antje du Bois-Pedain & Ulfrid Neumann eds., 2014).

[83] I assume here that HP1 is not satisfied if criminalizing all trespass would probably prevent no more harm than criminalizing harmful trespass alone. This is to assume that harmful trespass is already prohibited in the legal system in question and that this prohibition itself satisfies HP1, such that the crucial question is whether criminalizing harmless trespass would probably prevent any further harm. This assumption favors my opponents, so I do not apologize for it here.

[84] Or, where harmless trespass is already criminalized, the effect of its decriminalization. Some think that the fact that HP1 requires this kind of empirical calculation counts against it. I do not see why this should be so. Assessing the legitimacy of much (perhaps all) governmental activity requires that we consider its likely effects. Nor, when activities are ongoing, can we always risk finding out whether we are right about how things would change if the activity ceased. Would slashing income tax lead to an explosion of productive enterprise that enriches all or would it impoverish many? Would getting rid of the minefield on the border lead to hostilities or pave the way to better relations? Sometimes the risk involved in finding out is too great. This does not mean our calculations about what would happen are irrelevant to the legitimacy of our activities.

[85] It might be said in reply that harmless trespass is permissibly criminalized even if doing so prevents no harm and that this is enough to falsify HP1. But this is far from obvious. Imagine some group of people is harmlessly trespassing somewhere at night without risk of detection by the owner, and preventing them doing so will not prevent any harm. Would it be permissible for us to take preventive steps if those steps would be highly likely to harm the group members (and their family and friends) in ways we cannot fully control, including by imposing harms they do not deserve? I doubt it but cannot argue the point here.

criminalized only if φing is harmful or creates a risk of harm the creation of which is unjustified. Now, it is at least arguable that the harmless trespasser does create such a risk. One might object that if the precautions taken are sufficient, the risk may be tiny. But creating a tiny risk is unjustifiable if there is little or no reason in favor of the risk-creating act and if the magnitude of the potential harm is great. So while I focus in what follows on the instrumental principle HP1, it is worth noting that act-centered harm principles may also permit criminalization of wrongs that are not harmful.[86]

VIII.

Many think there are decisive objections to HP1. One such objection is the wrong kind of reasons objection, WO. According to this objection, we criminalize harmless wrongs such as A's trespass (at least in part) *because* they are wrongs. Yet, so the objection goes, HP1 implies that the *only* reason to criminalize harmless wrongdoing is to prevent further harmful acts (such as C's harmful trespass on D's land), and/or to prevent reactive harms—harms that are the result of individual reactions to changes in the law (such as E's becoming unable to enjoy her home if harmless trespass is decriminalized).[87] As Antony Duff puts it, HP1 "distorts the reasons for criminalising" these wrongs;[88] it "cannot capture the grounds that we surely have for criminalising several kinds of wrongful conduct."[89] In Arthur Ripstein's words, HP1 "cannot provide an adequate account of either the wrong I commit against you or the grounds for criminalising it."[90] And according to Hamish Stewart, "what justifies making rape a crime" if we endorse HP1 "is not that it is a *wrong to the victim*, but that it is *harmful to others* to allow it to occur." In Stewart's view, this "fails to take the individual victim seriously."[91] In short, even if HP1 permits criminalization of some harmless wrongs, it offers us a defective account of the reasons for doing so.

One might seek to reply to WO by again invoking DT. One might argue that HP1 is an instrumental, not a purposive, principle. Unlike HP5, HP1 says nothing about the aims that lawmakers can permissibly pursue, and thus cannot fairly be criticized for presenting a defective account of those aims. This reply appears to run up against two problems. First, it assumes that WO is a thesis about motivating reasons—that is, the facts that lawmakers take to count in favor of criminalization. Yet WO is also plausibly addressed to normative reasons—that is, to the facts that actually count in favor of criminalization, whether or not anyone ever takes these facts to count in

[86] I owe this point to Doug Husak and Andrew Simester.
[87] I borrow the term reactive harms from SIMESTER & VON HIRSCH, *supra* note 17, at 47–48.
[88] DUFF, ANSWERING, *supra* note 67, at 129 n.48.
[89] *Id.* at 135.
[90] Ripstein, *supra* note 77, at 218.
[91] Hamish Stewart, *The Limits of the Harm Principle*, 4 Crim. L. & Phil. 17 (2010), at 30 (emphasis in original).

favor of anything. If HP1 has implausible implications when it comes to normative reasons, this is cause to reject it.

Second, the reply assumes that there is no connection between HP5 and HP1—that the latter has no implications for the legitimate aims of those who criminalize. This can be doubted. As Simester and von Hirsch point out, HP1 is most plausibly viewed as a positive constraint.[92] Rather than merely eliminating (normative) reasons not to criminalize, the prevention of harm is most plausibly viewed as a (normative) reason to criminalize. And if the clumsiness argument is correct, there are (normative) reasons not to criminalize—deriving from the harm done by criminalization—which can only be defeated by harm prevention. Now, on a plausible view of what it means for an action to be justified, one is justified in φing only if there is undefeated reason to φ *and* one φs for an undefeated reason.[93] It seems to follow that if HP1 is true, lawmakers are justified in criminalizing only if they are likely to prevent harm thereby and criminalize in order to prevent harm. Yet this latter demand is the very demand imposed by HP5. HP1 thus seems to entail HP5.[94] If WO is an objection to HP5, it appears that it is also an objection to HP1.

In fact, whatever its merits as an objection to HP5, WO fails as an objection to HP1. To see this, return to CT. As Section V shows, HP1 is compatible with a number of other constraints. In particular, HP1 is compatible with 9):

> 9) that φing is morally wrong is a necessary condition of the permissibility of criminalizing φing.

Why might one endorse 9)? Simester and von Hirsch argue that because criminal conviction expresses public condemnation of the convicted and because public condemnation implies that *p* has committed a moral wrong, to convict *p*, who has committed no such wrong, is a form of public defamation.[95] To avoid defaming its citizens, the state must thus criminalize only moral wrongs. This line of thought implies that 9) is a negative constraint—if 9) is conformed to, some reasons not to convict will be canceled in the case of those convicted. Other lines of thought suggest that 9) is a positive principle. It may be that the state has reason to express disapproval of (at least some) moral wrongdoing and that this disapproval is expressed adequately only if (at least some) moral wrongs are criminalized. It may also be

[92] Simester & von Hirsch, *supra* note 17, at 35.

[93] *See* Gardner, *supra* note 64, at 91–120.

[94] There are several assumptions on which this alleged entailment depends, including a) the truth of the Gardnerian view of justification, and b) the truth of the claim that there are reasons not to criminalize, which can be defeated only by reasons to prevent harm. If these assumptions are false, HP1 does not entail HP5, and objections to HP5 are not necessarily objections to HP1. I concede the truth of the assumptions here and argue below that WO can be rebutted even with the concession.

[95] Simester & von Hirsch, *supra* note 17, at 9, 19–20. In principle, the state could criminalize φing but never prosecute any offender. But it is plausible to think this would itself be impermissible—that the state should not pass laws it has no intention of enforcing.

https://doi.org/10.1017/S135232521500004X Published online by Cambridge University Press

that the state has reason to give (at least some) moral wrongdoers what they deserve and that even if this could in principle be done without criminalization, (at least some) wrongs must be criminalized if condemnation of those wrongdoers is to adhere to demands of the rule of law. I here acknowledge parenthetically that some doubt that the state has any reason to disapprove of, or punish, all moral wrongdoing. If we share these doubts, we may add 10) to 9):

10) that φing is the business of all citizens is a necessary condition of the permissibility of criminalizing φing.[96]

Whether restricted in this way or not, if there is reason to criminalize moral wrongs, then 9) is a positive principle, and if we accept something like Simester and von Hirsch's argument, it is a positive constraint.

What relevance does this have for WO? Remember that HP1, like any harm principle, cannot plausibly be thought to be a sufficient condition of permissible criminalization. If we endorse 9), one additional necessary condition is that the conduct criminalized is morally wrong. On this view, criminalization of φing is impermissible unless i) φing is morally wrong, and ii) criminalization of φing is likely to prevent harm. Even though HP1 is a positive constraint and thus identifies a reason to criminalize, that reason is defeated unless 9) is satisfied; only when i) and ii) obtain is there undefeated reason to criminalize φing.[97]

Now recall that according to WO, HP1 implies that the only reason to criminalize harmless wrongdoing is the fact that this will probably prevent harm. As shown above, this can be interpreted as a claim about either normative or motivating reasons. Take normative reasons first. Does HP1 imply that the only fact that counts in favor of criminalizing a harmless wrong is the prevention of harm? No. As just shown, HP1 implies that the prevention of harm is a normative reason to criminalize φing but that this reason is itself insufficient to defeat all reasons not to criminalize. And HP1 is compatible with claiming that the fact φing is a harmless wrong itself generates a reason to criminalize φing, which reason is also insufficient itself to defeat all countervailing reasons.

To see the point another way, consider 11):

11) that φing will prevent moral wrongdoing is a reason to φ.[98]

[96] Some would add a further condition: φing must be a wrong that is the business of all citizens *and* one that must be criminalized if the citizenry are to take certain values seriously. For this suggestion, *see* R.A. Duff & S.E Marshall, *Public and Private Wrongs*, *in* Essays in Criminal Law in Honour of Sir Gerald Gordon (James Chalmers, Fiona Leverick & Lindsay Farmer eds., 2010).

[97] And there will be further conditions that must also be met. I do not always add this qualification.

[98] Some would wish to add that the wrongdoing must be the business of the actor. This qualification can be read into the text without affecting the argument.

https://doi.org/10.1017/S135232521500004X Published online by Cambridge University Press

If 11) is sound and lawmakers are faced with a currently unregulated wrong, *w*, they have reason to regulate *w* if doing so will reduce its incidence. If *w* is a serious wrong and occurs often, that reason may be a strong one, even if *w* is harmless. Now imagine that the incidence of *w* would be most drastically reduced by criminalization. If HP1 is correct, criminalization is justified only if criminalizing *w* will probably prevent harm.[99] But it is implausible to think that the reason given by 11) disappears merely because the regulatory tool under consideration is criminalization. Nor does HP1 imply disappearance. HP1 simply adds that the reason mentioned in 11) is defeated by reasons not to criminalize, unless and until it can be shown that criminalization is also likely to prevent harm. So HP1 does not imply that the only normative reason to criminalize harmless wrongdoing is harm prevention. Conceived in terms of normative reasons, WO thus fails.

Now consider motivating reasons. Does HP1 imply that criminalization is permissible only if lawmakers act in order to prevent harm? No. We can concede that criminalization is justified only if lawmakers act for an undefeated reason. If we accept HP1 and 9), there is undefeated reason to criminalize only if, inter alia, i) φing is morally wrong, and ii) criminalization of φing will probably prevent harm. Both i) and ii) are plausibly thought of as providing reasons to criminalize, yet each reason is undefeated only when the other exists: that criminalization of φing will probably prevent harm provides an undefeated reason to criminalize only if φing is also morally wrongful; that φing is morally wrongful provides an undefeated reason to criminalize only if criminalization of φing will probably prevent harm.

Now, if justification requires action for any undefeated reason, then when i) and ii) obtain, lawmakers can act for either reason. It follows that lawmakers whose actions are justified can act for i): where φing is a harmless wrong, lawmakers can criminalize that wrong because it is wrongful. HP1 thus does not entail HP5. One might object to this as follows. If neither i) nor ii) can itself defeat the reasons not to criminalize φing, then only the conjunction of i) and ii) amounts to an undefeated reason to criminalize. So lawmakers whose actions are justified must act for the composite reason that φing is morally wrongful and that criminalizing φing will probably prevent harm. But even if this is correct, HP1 still does not imply, as WO claims it does, that lawmakers must criminalize only in order to prevent harm. In fact, it implies that lawmakers must also criminalize φing for the reason that φing is wrong. Conceived in terms of motivating reasons, WO thus fails.

There is a more general lesson to be learned from this discussion. That lesson is that HP1 is a more modest principle than many of its critics suppose. *Pace* Duff and Stewart, HP1 does not tell us that the only good reason to criminalize—normative or motivating—is the prevention of harm.[100] As just shown, if we think the wrongfulness of φing is a good reason to criminalize

[99] Again, here I assume that x is criminalization.

[100] Duff writes that he is concerned with "the question of what can count in principle as a good reason for criminalising a certain kind of conduct" (Duff, Answering, *supra* note 67,

https://doi.org/10.1017/S135232521500004X Published online by Cambridge University Press

φing, we need have no quarrel with HP1. Nor does HP1 misdescribe the wrongs involved in cases of battery, trespass, or rape. *Pace* Ripstein, HP1 gives no account at all of what makes any action wrongful. It certainly does not say that battery, trespass, and rape are wrongs because of the harm they cause.[101] As Section IV shows, some actions are wrongful partly because they are harmful; others are wrongful independent of harm. But to repeat, HP1 has nothing to say about any of this. Of any wrong, HP1 merely says that whatever reasons we have to criminalize that wrong are insufficient to justify criminalization unless criminalizing the wrong is also likely to prevent harm.

IX.

The argument of the previous section helps us rebut a further objection to HP1. That objection begins with the thought that it is impermissible to treat anyone as a mere means of achieving one's ends or the ends of others. Some worry that this means principle is in conflict with HP1. Why? Because HP1 suggests that it is sometimes permissible to criminalize the actions of A in order to prevent B from suffering harm at the hands of C.[102] Yet if A's actions are criminalized for this reason, A appears to be treated as a mere means of achieving harm-preventive ends. As Andrew Ashworth and Lucia Zedner put the point:

> if the individual is not to be treated merely as a means to the ends of others but as an end in her own right, it cannot be justifiable to impose public censure and punishment on her in order to prevent a harm that might . . . be done, were this particular offense not to be a crime.[103]

This is the mere means objection, MO.[104]

MO is a good objection only if those who criminalize in order to prevent harm treat people as a mere means and only if treating people as a mere means is impermissible. The word *mere* is not expendable in either part of

at 135) and takes the harm principle to state that the only such reason is to prevent harm. Stewart writes of a "refusal of the harm principle to recognize that, sometimes, a pure violation of rights is a good reason to criminalize"; Stewart, *supra* note 91, at 19.

[101] John Gardner replies to Ripstein in similar vein: *see* GARDNER, *supra* note 64, at 241–243.

[102] As in the previous section, I assume here that HP1 is a positive principle, such that the fact that φing will probably prevent harm is a reason to criminalize. If that is so, then in cases where criminalization is justified, that reason is (part of) an undefeated reason, and lawmakers can permissibly act for it.

[103] Andrew Ashworth & Lucia Zedner, *Prevention and Criminalization: Justifications and Limits*, 15 New Crim. L. Rev. 542 (2012) at 551.

[104] Ripstein makes a similar objection when he writes that a principle such as HP1 will inevitably "run afoul of the liberal ideal of individual responsibility, which says that you cannot be prohibited from doing something that does not interfere with anyone else, simply because prohibiting you from doing it makes you or someone else less likely to commit some *other* genuine crime." *See* Ripstein, *supra* note 77, at 228 (emphasis in original).

this conjunction. It is not generally objectionable to treat people as a means of achieving our ends. As G.A. Cohen puts it:

> *Of course* I treat the ticket-seller as a means when I hand him the money and thereby get him to hand me my ticket. For I interact with him only *because* he is my means of getting a ticket. After all, I undoubtedly treat the ticket machine as a means when I put my money in it, and there is plainly something in common between how I treat or regard the machine's action, what its place within my purposes is, and what the ticket-seller's place within my purposes is. The action that I induce in each case serves as a means to my purposes, and that is why I induce it.[105]

But the fact that I treat the ticket-seller as a means does not show that I treat him as a *mere* means, and it is only if I treat him as a mere means that I act impermissibly. As Cohen says, I do not treat the ticket-seller as a mere means "*provided that* I at the same time honour his status as an independent centre of value, as an originator of projects that demand my respect."[106] One way to show this respect is to acknowledge that the ticket-seller is the source of certain restrictions on how I can permissibly act. Perhaps if I simply punched the ticket-seller in the face and took my ticket, I could properly be accused of treating him as a mere means. But if I simply pay for my ticket, I cannot be accused of acting impermissibly.

Cohen's point is of relevance here for the following reason. In the criminalization context, one plausible constraint on how lawmakers can permissibly act is 9).[107] As shown in the previous section, Simester and von Hirsch argue that φing must be morally wrong if it is to be criminalized, because this is the only way to avoid public defamation of those convicted of crimes. One might add that because conviction is frequently followed by punishment, we should also accept 12):[108]

12) that those who φ deserve punishment is a necessary condition of the permissibility of criminalizing φing.[109]

One might endorse 12) because one believes that p is wronged if p does not deserve punishment but is punished nonetheless. No doubt there are further principles we might add to the list. As CT shows us, many such principles, including 9) and 12), are compatible with HP1. Let us thus imagine

[105] Gerald Cohen, Self-Ownership, Freedom, and Equality (1995), at 239 (emphasis in original).
[106] *Id.* (emphasis in original).
[107] According to 9), that φing is morally wrong is a necessary condition of the permissibility of criminalizing φing.
[108] 12) is endorsed by Douglas Husak; *see* Husak, *supra* note 28, at 82–91.
[109] A defender of 12) need not claim that φers deserve punishment prior to the criminalization of φing; one may deserve punishment for φing partly because it is a crime. Once traffic offenses are created, others' reliance on those laws may make my offending acts especially dangerous. Those acts may then be deserving of punishment, even though they would not have been had the traffic offenses never come into existence.

lawmakers who endorse all three principles.[110] They criminalize φing in order to prevent harm that would occur if φing were not criminalized. But they criminalize φing only on the condition that φing is morally wrongful and that φers deserve punishment. Are those who φ treated as mere means? I think not. The lawmakers in question clearly do acknowledge that φers are the source of restrictions on how they can permissibility act: φers must not be publicly defamed or subjected to undeserved punishment. True, by criminalizing φing, lawmakers treat those who φ as a means of achieving their harm-preventive ends. But they do not treat them as *mere* means. So they do not violate the means principle.

One might reply by saying that the means principle cannot be satisfied so easily. If I punch the ticket-seller in the face, I cannot evade the accusation of having treated him as a mere means simply by pointing out that I refused to kill him (even though, let us assume, that would have been easier for me). Imagine I refuse to kill the ticket-seller because I acknowledge that he is the source of a moral restriction on killing. I surely still act impermissibly by punching him in order to get my ticket. There are difficult questions here about what is required in order to act compatibly with the means principle. But we do not need to answer these questions to show that the reply fails. In the ticket-seller example, the customer fails to acknowledge a moral constraint we intuitively find plausible (he takes himself to be permitted to assault others to further trivial interests of his own). This is what motivates the thought that he treats the ticket-seller impermissibly. To endorse HP1 is not like this. If we endorse 9) and 12), to add HP1 is to add a moral constraint to the picture, not to fail to recognize one. Not only must φing be morally wrongful and not only must φers deserve punishment, but criminalization of φing must also prevent harm. This only offers people greater protection against the criminal law.

It might be said in further reply that we miss the point by focusing on criminalization itself. Perhaps the real objection to HP1 is that if φing is criminalized in order to prevent harm, some increments of punishment will be imposed on offenders for harm-preventive ends, increments that would not have been imposed had HP1 been rejected. Those on whom these increments are imposed are treated as mere means. This objection also fails. HP1 does not offer or imply any theory of punishment; in particular, that it is permissible for lawmaking officials to criminalize in order to prevent harm does not imply that it is permissible for law-applying officials to impose a certain quantum of punishment to prevent harm. To criminalize φing is to make those who φ liable to particular burdens (such as criminal conviction) if particular conditions are met (such as proof of guilt beyond reasonable doubt in a criminal trial). It is one thing to ask why those who commit tokens

[110] Whether they are right to do so is a separate question. The point here is not to argue for 9) or 12). It is rather to show that lawmakers committed to HP1 need not treat people as mere means.

https://doi.org/10.1017/S135232521500004X Published online by Cambridge University Press

of any act-type should be liable to such burdens. It is another entirely to ask when those burdens should be imposed on those liable to them and how much of a particular burden anyone should suffer.[111] HP1 has nothing to say in answer to this latter question. MO is thus no objection to HP1.[112]

X.

Many press a further objection against HP1. According to this objection, HP1 should be rejected because it amounts to an intolerably weak constraint. According to Douglas Husak, HP1 "has the potential to expand the scope of the criminal law exponentially,"[113] and according to Antony Duff, it "can do little work in limiting the expansion of the criminal law."[114] To use an example of Hamish Stewart's: if enough people are traumatized by the prospect of people in their society consuming carrot juice, and if this trauma makes it difficult for them to pursue valuable projects and goals, criminalization of said consumption may satisfy HP1.[115] As Victor Tadros points out, even criminalization of homosexual sex—and indeed, of heterosexual sex—seems likely to prevent some harm by preventing the formation of some sexual relationships, some of which would have been harmful.[116] These examples suggest not only that the limitations imposed by HP1 are, in Meir Dan-Cohen's words, "very feeble,"[117] but that HP1 permits the criminalization of actions that—we would all accept—cannot permissibly be criminalized. This is the overinclusiveness objection, OO.

A quick response to OO relies on CT. We have seen that HP1 is compatible with 9), 10), and 12). If we endorse any of these principles, it is impermissible to criminalize the consumption of carrot juice, homosexual sex, or heterosexual sex. These act-types are not wrongs. They are not the business of all citizens. One does not deserve punishment for committing them. Signing up to HP1 thus does not commit one to the view that criminalizing such act-types is permissible.

[111] Most obviously, if the burdens in question carry a degree of moral censure, it may be that fairness to offenders requires that judges impose only a quantum of punishment that reflects offenders' moral blameworthiness. HP1 is perfectly consistent with this thought.

[112] The objection made in this paragraph is the flip side of one of Patrick Devlin's objections to the harm principle. Devlin claims that if moral wrongness is relevant to the quantum of punishment we should impose for φing, it must also be relevant to whether we should criminalize φing in the first place; see PATRICK DEVLIN, THE ENFORCEMENT OF MORALS (1965), at 130–131. The present objector claims that if harm prevention is relevant to whether we should criminalize φing, it must also be relevant to the quantum of punishment we should impose for φing. My response is similar to that pressed by a number of writers against Devlin; *see, e.g.*, H.L.A. HART, PUNISHMENT AND RESPONSIBILITY (2d ed. 2008), at 8–13; FEINBERG, HARMLESS, *supra* note 25, at 144–151.

[113] HUSAK, *supra* note 28, at 72.

[114] DUFF, ANSWERING, *supra* note 67, at 138.

[115] Stewart, *supra* note 91, at 28–29.

[116] Tadros, *Harm, Sovereignty*, *supra* note 49, at 50.

[117] MEIR DAN-COHEN, HARMFUL THOUGHTS (2002), at 152.

https://doi.org/10.1017/S135232521500004X Published online by Cambridge University Press

This reply is too quick. If we are forced to rely on 9), 10), or 12) whenever we are faced with an unpalatable criminal offense, HP1 appears to do no work in setting the limits of the criminal law. If we criminalized picking daisies, some harm would probably be prevented—some back injuries would not occur, some people would stay away from flowers to which they suffer severe allergic reactions. If HP1 permits criminalization in such a case, there can be very little that HP1 rules out.

These remarks assume that HP1 requires simply that criminalization will probably prevent some harm. But as mentioned in Section IV, a complete statement of HP1 would contain additional conditions. First, it would say something about how criminalization compares with alternative means by which some harm might be prevented. Second, it would say something about how criminalizing and preventing the harm compares with doing nothing to prevent the harm. Neither of these comparisons is straightforward.

As to the first, there will often be other ways to prevent harm that could be prevented via criminalization. In some cases, we may be morally required to use these other means. This suggests we should endorse what I call the *alternative means condition* (AMC). One way to interpret AMC is as a test of necessity: if there is any alternative way of preventing the harm in question, criminalization is unnecessary and is ruled out by AMC. This cannot be right. Imagine we can prevent some harm either by sending a section of the populace to a gulag or by criminalizing φing. The existence of the first alternative does not make criminalization impermissible. We must also take into account the fact that the gulag is morally worse, all things considered, not only because there are additional reasons against using the gulag (such as violations of the rule of law), but also because there are additional reasons for criminalization (such as the expressive value of censuring φers if φing is a serious wrong).[118]

Now imagine instead that we could also prevent the harm by funding an advertising campaign. Imagine there are several additional reasons for the campaign that carry substantial moral weight (it does much less harm, it is much less costly), and only one additional reason against criminalization (the expressive value of censuring φers). All else being equal, it is plausible to think that criminalization is impermissible.

These examples suggest a better interpretation of AMC. To work out whether AMC is satisfied, we should first compile a list of the means by which some harm could probably be prevented. Assuming criminalization is on the list, such that HP1 is potentially satisfied, we should then compare criminalization to each alternative, considering the pros and cons of criminalization and the pros and cons of the alternative. If any alternative is

[118] This point is well made by Husak: *see* Douglas Husak, *The Criminal Law as Last Resort*, 24 Oxford J. Legal Stud. 207 (2004).

morally better than criminalization, all things considered, then criminalization does not satisfy AMC. If there is no such alternative, AMC is satisfied.[119]

It might be argued that on this interpretation, anything that satisfies AMC will be morally justified, all things considered. AMC then collapses into all-things-considered moral justification. This is not the case. Even if all the alternatives are morally worse than criminalization, criminalization may remain unjustified; despite being the least bad alternative, it may be too bad, all things considered. Imagine that the only way to prevent some harmful religious practice is either i) to permit private contractors legally to round up all believers and lock them in individual cells, or ii) to criminalize the harmful religious practice and devote enormous resources to enforcement of the new offense. It is plausible to think that both options would be unjustified, all things considered, even though criminalization may well satisfy AMC.

This leads us to the second comparison: even if criminalization satisfies AMC, we must also compare criminalization with doing nothing to prevent the harm. This condition is sometimes referred to as the *proportionality condition* (PC): the harm brought about by criminalization must not be disproportionate to the harm it prevents. If it is, we should not prevent the harm at all. While there are reasons to hesitate in using this label, I adopt it here for convenience.[120] One might think that PC is satisfied if and only if the harm we expect to be prevented by criminalization is greater than the harm we expect it to bring about. This is too simplistic. We must also consider the likelihood that the harms will be prevented or brought about. And we must work out how much moral weight to give to these harms.[121]

There are difficult questions to answer here, and I can do no more than scratch the surface. Imagine a case in which criminalization would probably prevent a total amount of harm, H_1, but would also probably bring about a lesser amount of harm, H_2. Imagine the two probabilities are the same. There are several reasons that it might be disproportionate to criminalize. One is that H_2 would be harm done by state officials whereas H_1 would merely be harm that those officials allowed. Another is that at least some of H_2 would be harm that state officials intended,[122] whereas H_1 would at

[119] Although I focus here on noncriminal alternatives, there is also the question of alternative criminal means of prevention. It is plausible to think that AMC also applies here. If some harm would probably be prevented by criminalizing φing or by criminalizing μing, it is plausible to think that it is permissible to criminalize φing only if criminalizing μing is not morally better, all things considered.

[120] One is that the term *proportionality* refers in some legal contexts to a set of conditions including a version of AMC.

[121] It might be said that we must also consider the other goods (and bads) that criminalization will bring about. It is true that the all-things-considered permissibility of criminalization requires reference to these things. But this is a separate question. If HP1 is sound, only the prevention of harm is capable of defeating the reasons not to criminalize given by the clumsily harmful nature of criminalization. PC is thus best interpreted as requiring consideration of whether this in fact occurs—whether the harm brought about by criminalization is disproportionate to the harm prevented by it.

[122] What Section II calls *internal punitive harms* are the most obvious candidate.

https://doi.org/10.1017/S135232521500004X Published online by Cambridge University Press

most be harm that those officials foresaw. If it is worse to do harm than it is to allow it or if intentional harming is worse than harming as a side effect, it may be worse to bring about the smaller H_2 even though one would then fail to prevent the larger H_1.[123] Criminalization would then violate PC.

Some, of course, would deny one or both of the moral claims on which this conclusion depends, or would accept those claims as applied to individuals but deny that they apply to the state.[124] Some would also claim that if more of H_2 would be harm done to those who are morally liable to be harmed, criminalization may not be disproportionate even if H_2 is larger than H_1. The fact that criminalization will likely prevent more harm than it causes may thus be neither necessary nor sufficient to satisfy PC.

There is clearly more to say here. But enough, I think, has been said to offer the beginnings of a response to OO. Once we add AMC and PC to our partial statement of HP1, we see that the limits that HP1 imposes on the criminal law are not the feeble limits they appear to be at first sight. Many offenses that would likely prevent some harm will fail to satisfy HP1, because there will be alternative ways to prevent the harm that are morally better or because it would be disproportionate to prevent it. True, the application of HP1 is now more complicated than it first appeared. But it is no objection to a principle that its application is complicated, so there is no objection here to HP1.

XI.

Victor Tadros argues that HP1 is paradoxical.[125] According to Tadros, HP1 implies 13):

13) it is wrong to prevent *p* from *φ*ing unless preventing *p* from *φ*ing prevents harm.

So if A harmlessly prevents B from *φ*ing and does not prevent harm thereby, A acts wrongly. But HP1 also seems to tell us that it is wrong to prevent A's wrong: if A's act was harmless, preventing it, let us assume, will not prevent any harm. Tadros thinks this paradoxical: HP1 implies both that A's preventive act is wrong, and that it would be wrong to prevent it.[126]

Our discussion to this point shows that the alleged paradox is illusory. 13) can be restated as follows:

13a) it is a necessary condition of permissibly preventing *p* *φ*ing that preventing *p* *φ*ing prevents some harm.

[123] For extended discussion of these points in a different context, *see* David Rodin, *Justifying Harm*, 122 Ethics 74 (2011).

[124] For discussion, *see* David Enoch, *Intending, Foreseeing, and the State*, 13 *Legal Theory* 69 (2007); Adam Hosein, *Doing, Allowing, and the State*, 33 *Law & Phil.* 235 (2014).

[125] This is one of several difficulties raised by Tadros for HP1, not all of which I can discuss here. *See* Tadros, *Harm, Sovereignty*, *supra* note 49.

[126] *Id.* at 58.

https://doi.org/10.1017/S135232521500004X Published online by Cambridge University Press

This shows us that in 13) what I call x is all acts of prevention. Yet Section II shows that there are many other candidates for x and that at least one argument for HP1—which I called the clumsiness argument—supports another. If that argument is sound, it is 14), and not 13), which is implied by HP1:

14) it is wrong to use tools that are clumsily harmful to achieve some objective (including a harm-preventive objective) unless doing so prevents some harm and AMC and PC are satisfied.[127]

Is 14) paradoxical? Certainly not in the same way as 13). 14) says nothing about the wrongness of harmless acts of prevention. 14) does not even say anything about the wrongness of many acts of prevention that are harmful. It does, of course, identify a candidate wrong. Does 14) imply that this wrong cannot be prevented without the preventer also acting wrongly? No. There is nothing in 14) to say that it would be wrong to prevent the wrong mentioned in 14) either by harmless means or by means that are harmful but not clumsy in the relevant sense. 14) does imply that if *p* pursues an objective using a tool that is clumsily harmful and does not prevent harm thereby (or does but violates AMC or PC), it would be wrong for *q* to use *another* such tool to prevent *p*'s action if this does not prevent harm (or violates AMC or PC). There is, however, nothing paradoxical about that.

[127] According to AMC, it is a necessary condition of permissibly criminalizing *φ*ing that there is no alternative means of preventing harm that is morally better, all things considered. According to PC, it is a necessary condition of permissibly criminalizing *φ*ing that the harm brought about by criminalization is not disproportionate to the harm it prevents.

https://doi.org/10.1017/S135232521500004X Published online by Cambridge University Press