What is the Explanatory Role of the Harm Principle (and Other Mid-Level Normative Principles)?[1]

Collis Tahzib

Abstract: For many liberals, the harm principle plays an important role within the liberal order of explanation. It explains—or at least helps to explain—what rights and duties we have. But on the view Joseph Raz develops in *The Morality of Freedom*, the harm principle cannot do this sort of explanatory work: it cannot help us to determine the contours of individual rights and duties because judgements about harm are themselves parasitic on prior judgments about the contours of individual rights and duties. In this paper, I explore these sorts of issues about the explanatory role of the harm principle. First, I suggest that the main ways of thinking about this role can be helpfully organized within a two-by-two grid according to whether or not the definition of "harm" (i) depends on other moral and political commitments and (ii) is systematic. Second, I argue in favour of a certain quadrant of this grid, the main upshot of which is that the harm principle does not itself explain individual rights and duties but is instead shorthand for a systematic normative theory. Third, I generalize the discussion by suggesting that this grid in general—and the favoured quadrant in particular—can be used to illuminate the explanatory role played by other important and influential principles within moral and political philosophy, such as the principle of public justification. Finally, I discuss an intriguing upshot of this argument: we should dispense with much mid-level normative theorizing and the central concepts it cites such as "harm" and "reasonableness" in favour of a more direct inquiry into the relative plausibility of different systematic theories of morality and justice.

1. A Razian Introduction

Many leading liberal political philosophers, such as John Rawls and Ronald Dworkin, affirm the principle of state neutrality.[2] On this view, the state should not impose or even promote any particular conception of the good life or human flourishing. It should instead restrict itself to maintaining a fair framework of rights and opportunities within which all citizens can pursue their own beliefs about what constitutes a good life. By contrast, perfectionist political philosophers—foremost among whom was Joseph Raz—argue that the state can and should

[1] For very helpful comments and discussions, I thank [omitted].

[2] See Dworkin, R., *Sovereign Virtue: The Theory and Practice of Equality* (Cambridge, MA: Harvard University Press, 2000); Rawls, J., *Political Liberalism* (New York: Columbia University Press, 2005). For other articulations of liberal neutrality, see, e.g., Gaus, G., "Liberal Neutrality: A Compelling and Radical Principle", in S. Wall and G. Klosko (eds), *Perfectionism and Neutrality: Essays in Liberal Theory* (Oxford: Rowman & Littlefield, 2003), pp. 137-65; Quong, J., *Liberalism without Perfection* (Oxford: Oxford University Press, 2011); Pallikkathayil, J., "Neither Perfectionism nor Political Liberalism", *Philosophy & Public Affairs* 44 (2016), pp. 171-96.

implement laws and policies designed to encourage citizens to lead flourishing lives, such as public funding of the arts and taxes on drugs and gambling.[3]

An important objection to perfectionist accounts of political morality is that they violate the harm principle. The harm principle says that the state may interfere with the choices of citizens only if those choices harm or risk harming others. In John Stuart Mill's words, "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".[4] Harmless actions, that is, are none of the state's business. Yet perfectionists hold that the state may legitimately interfere with the choices of citizens for the purpose of steering them away from unworthy pursuits and towards more edifying avenues, even when those choices do not harm or risk harming third parties.

Some perfectionists respond to this objection by simply denying that the harm principle is a sound principle of political morality. In *The Morality of Freedom*, however, Raz offers an ingenious response of a more conciliatory kind. He argues that (1) an action is harmful just in case it violates positive or negative duties and that (2) the state has a positive perfectionist duty to enact laws and policies designed to promote the good life and human flourishing and thus that (3) the activities of a perfectionist state are compatible with the harm principle. As Raz puts this argument:

> The government has an obligation to create an environment providing individuals with an adequate range of [valuable] options and the opportunities to choose them. The duty arises out of people's interest in having a valuable autonomous life. Its violation will harm those it is meant to benefit. Therefore its fulfillment is consistent with the harm principle.[5]

For our purposes here, premise (1) is crucial. As Raz explains:

> One can harm another by denying him what is due to him. This is obscured by the common misconception which confines harming a person to acting in a way the result of which is that that person is worse off after the action than he was before. While such actions do indeed harm, so do acts or omissions the result of which is that a person is worse off after them than he *should* then be. One harms another by failing in one's duty to him, even though this is a duty to improve his situation and the failure does not leave him worse off than he was before.[6]

---

[3] For articulations of perfectionism, see, e.g., Raz, J., *The Morality of Freedom* (Oxford: Clarendon Press, 1986); Sher, G., *Beyond Neutrality: Perfectionism and Politics* (Cambridge: Cambridge University Press, 1997); Wall, S., *Liberalism, Perfectionism, and Restraint* (Cambridge: Cambridge University Press, 1998); Tahzib, C., *A Perfectionist Theory of Justice* (Oxford: Oxford University Press, 2022).
[4] Mill, J. S., *On Liberty* (London: Penguin, 1985), p. 68.
[5] Ibid., pp. 417-8.
[6] Raz, J., *The Morality of Freedom*, p. 416 (emphasis added).

Raz further states that "since 'causing harm' entails by its very meaning that the action is prima facie wrong, it is a normative concept acquiring its specific meaning from the moral theory within which it is embedded. Without such a connection to a moral theory the harm principle is a formal principle lacking specific concrete content and leading to no policy conclusions".[7]

However, a crucial implication of Raz's argument—yet one on which he does not much comment—is that it makes the harm principle explanatorily redundant. For many liberals, the harm principle plays an important role within the liberal order of explanation. It explains—or at least helps to explain—what rights and duties we have. But on Raz's approach the harm principle cannot do this sort of explanatory work: it cannot help us to determine the contours of individual rights and duties because judgements about harm are themselves parasitic on prior judgments about the contours of individual rights and duties. On Raz's approach, that is, the harm principle is conceptually downstream: it comes later in the day, only once we have figured out the deep and difficult questions about the correct "moral theory" (or theory of justice) that specifies "what is due" to each person. Indeed, if Raz's argument is sound, we may well begin to wonder whether we should dispense with the harm principle altogether. What is the point of invoking the harm principle, if it does no real explanatory work?

Rather surprisingly, there has been no explicit and focused discussion of these sorts of issues about what Raz's perfectionist reinterpretation of the harm principle implies for the explanatory role of the harm principle. In this paper, I first suggest that the main ways of thinking about this role can be helpfully organized within a two-by-two grid according to whether or not the definition of "harm" (i) depends on other moral and political commitments and (ii) is systematic (Section 2). I then argue in favour of a certain quadrant of this grid, the main upshot of which is that the harm principle does not itself explain individual rights and duties but is instead shorthand for a systematic normative theory (Section 3). Next I generalize the discussion by suggesting that this grid in general—and the favoured quadrant in particular—can be used to illuminate the explanatory role played by other important and influential principles within moral and political philosophy, such as the principle of public justification (Section 4). Finally, I discuss an intriguing upshot of this argument: we should dispense with much mid-level normative theorizing and the central concepts it cites such as "harm" and "reasonableness" in favour of a more direct inquiry into the relative plausibility of different systematic theories of morality and justice (Section 5).

2. A Grid of Explanatory Roles for the Harm Principle

The main ways of thinking about the explanatory role played by the harm principle can, I propose, be helpfully understood as occupying positions within a two-by-two grid. Where a given account of the harm principle falls within this grid depends on how it answers two questions about the definition of "harm":

[7] Ibid., p. 414.

1. *The Moralization Question*: Does the definition of harm depend on other moral and political commitments?

2. *The Systematicity Question*: Is the definition of harm systematic?

These questions are distinct and so they give rise to four possible views about the explanatory role of the harm principle. I explore these four views in more depth in Section 3, but for now I offer a brief characterization of each view and mention some of its representative defenders.[8] I should, however, issue the caveat that it is sometimes difficult to know exactly where to place contemporary proponents of the harm principle within this grid—in part precisely because they are not always fully clear and explicit about how they define harm—and so a certain degree of conjecture and approximation is involved here.

- *The Moralized-and-Systematic View:* On this view, harm is defined in terms of some systematic moral or political theory such as Kantianism, Rawlsianism, libertarianism and perfectionism. Jonathan Quong and Joseph Raz can be understood as adopting this sort of view.[9]

- *The Moralized-and-Unsystematic View:* On this view, harm is defined in terms of various piecemeal principles or maxims, such as the maxim that offence is not harmful or the maxim of *volenti non fit injuria*. Andrew Jason Cohen, Joel Feinberg, Gerald Gaus and Andrew Simester and Andreas von Hirsch can be understood as adopting this sort of view.[10]

- *The Non-moralized-and-Systematic View:* On this view, harm is defined in terms of some systematic non-moralized account of harm, such as the temporal comparative account, the counterfactual comparative account, the non-comparative account, or the hybrid account. Anna Folland can be understood as adopting this sort of view.[11]

---

[8] An interesting exegetical question on which I do not take a stand is that of where Mill himself falls within this grid. For an interpretation that falls most naturally in the moralized-and-systematic quadrant, see Crisp, R., *Routledge Guidebook to Mill on Utilitarianism* (London: Routledge, 1997), ch. 8; for the moralized-and-unsystematic quadrant, see Rees, J. C., "A Re-Reading of Mill on Liberty", *Political Studies* 8 (1960) pp. 113-29; for the non-moralized-and-unsystematic quadrant, see Mulnix, M., "Harm, Rights, and Liberty: Towards a Non-Normative Reading of Mill's Liberty Principle", *Journal of Moral Philosophy* 6 (2009), pp. 196-217. I am not aware of an interpretation of Mill that would fall most naturally in the non-moralized-and-systematic quadrant (e.g. by attributing to Mill a commitment to some systematic non-moralized account of harm).

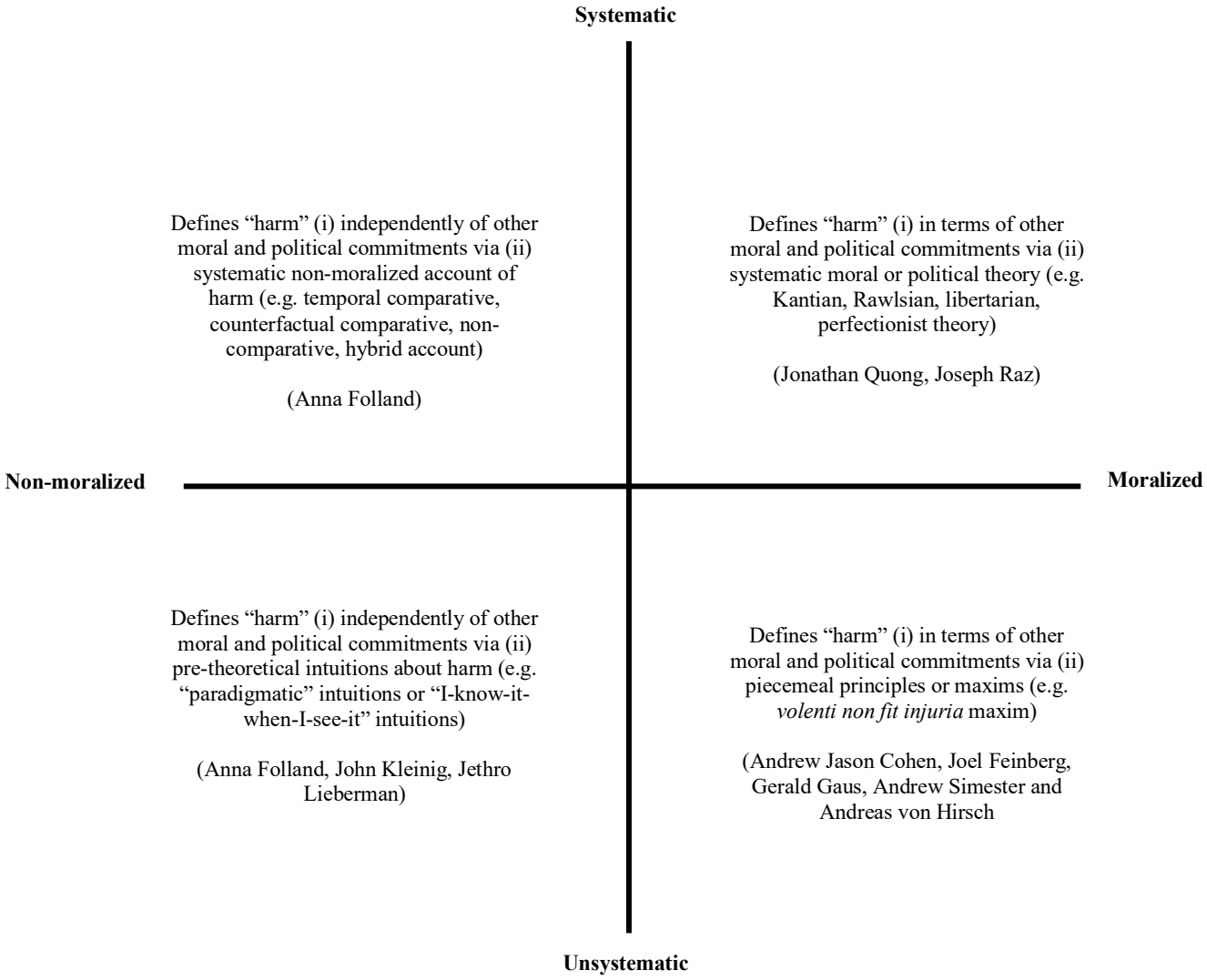
[9] See Raz, J., *The Morality of Freedom*, ch. 15; Quong, J., *Liberalism Without Perfection*, pp. 55-6.

[10] See Feinberg, J., *Harm to Others: The Moral Limits of the Criminal Law, Volume 1* (Oxford: Oxford University Press, 1984); Gaus, G., *Social Philosophy* (London: Routledge, 1999), ch. 8; Simester, A. and von Hirsch, A., *Crimes, Harms, and Wrongs: On the Principles of Criminalization* (Oxford: Hart Publishing, 2011); Cohen, Andrew J., *Toleration and Freedom from Harm: Liberalism Reconceived* (London: Routledge, 2018), esp. chs 3 and 6.

[11] See Folland, A., "The Harm Principle and the Nature of Harm", *Utilitas* 34 (2022), pp. 139-53.

- *The Non-moralized-and-Unsystematic View:* On this view, harm is defined in terms of our pre-theoretic intuitive judgements about harm in its natural-language sense. Folland (in some of her other arguments), John Kleinig and Jethro Lieberman can be understood as adopting this sort of view.[12]

We can represent this as follows:



3. In Defence of the Moralized-and-Systematic View

The explanatory role of the harm principle is, I think, most plausibly understood along the lines of the moralized-and-systematic view. To defend this position, I first set out the moralized-

[12] See ibid., pp. 150-1; Kleinig, J., "Crime and the Concept of Harm", *American Philosophical Quarterly* 15 (1978), pp. 27-36; Lieberman, J., *Liberalism Undressed* (Oxford: Oxford University Press, 2012).

and-systematic view in greater depth (Section 3.1). I then argue against the three alternative views found in the other quadrants of the two-by-two grid (Sections 3.2).

3.1 The Harm Principle as Moralized and Systematic

The moralized-and-systematic view holds that the definition of harm depends on other moral and political commitments (hence the *moralized*) and that these commitments take the form of a systematic moral or political theory such as Kantianism, Rawlsianism, libertarianism or perfectionism as opposed to a collection of piecemeal principles or maxims (hence the *systematic*).

An immediate and important implication of this view is that the harm principle is *redundant* (or at least very close to redundant) in the sense that it does not play any role (or at least it plays only a very minimal role) in explaining or justifying individual rights and duties.[13] Consider the familiar liberal right to absolute freedom of conscience—that is, the right to "absolute freedom of opinion and sentiment on all subjects, practical or speculative, scientific, moral, or theological".[14] What explains or justifies this right? What can we say in support of this right to illiberal individuals who would try to suppress freedom of conscience? As we will see, a common answer is we have a right to freedom of conscience because our thoughts and beliefs—what Mill calls "the inward domain of consciousness"—do not harm anyone else, where this claim about harmlessness is in turn justified either by consulting some systematic non-moralized account of harm (as per the view in Section 3.2.2) or by consulting our pre-theoretical intuitions about harm (as per the view in Section 3.2.3).[15]

By contrast, on the moralized-and-systematic view the right to freedom of conscience is not justified on the basis of any claims about harmfulness or harmlessness. Instead, this right is justified on some separate non-harm-based grounds, such as on the grounds of a certain conception of autonomy or self-ownership or respect for persons. The details of those non-harm-based grounds will vary according to the systematic moral or political theory within which the harm principle is embedded. One can then express or represent this justification using the harm principle: one can, if one wants, say that thought and belief are harmless and are to be protected from interference (though for reasons given in Section 5 I suspect that the language of harm is more trouble than it is worth and simply invites confusion and misunderstanding). But on this picture facts about the harmfulness or harmlessness of thought and belief do not themselves do play any role in explaining or justifying the right to freedom of conscience. As David Brink explains, "harm, on this reading, is the conclusion of an

[13] I add the qualifications in parentheses because (as I go on to explain later in this section) I think some kind of fidelity to the underlying motivations of the harm principle imposes a constraint on the sort of first-order views to which it can be wedded. To this extent, the harm principle is not entirely redundant. Still, the relevant constraint is very weak, and so the harm principle is very close to being redundant in the sense that it plays only a very minimal role in explaining or justifying individual rights and duties.
[14] Mill, J. S., *On Liberty*, p. 71.
[15] Ibid.

argument about whether it is permissible to interfere with an agent's liberty, not an independent premise in that argument" (though he immediately adds that this reading should be rejected, at least as an interpretation of Mill, "precisely because Mill appeals to harm as a reason for interference").[16]

This point about the redundancy of the harm principle can be put more generally. Suppose someone says that we should (or should not) interfere with some course of action because that course of action is harmful (or harmless). When we ask *why* that course of action is harmful (or harmless), the answer given by a proponent of the moralized-and-systematic view will cite some prior normative theory—or, in Raz's words, some "moral theory" that specifies "what is due" to each person. But then doesn't this prior normative theory do all the explanatory work here? What does the fact of harmfulness or harmlessness add to this theory? Once the prior normative theory about what is due to each person is on the table, hasn't the explanation already been given for why we should (or should not) interfere with some course of action? That a course of action is harmful (or harmless) it appears, is not what *makes* it the case that ought (or ought not) interfere with it but is instead merely a *by-product* of the reasons in virtue of which we ought (or ought not) interfere with it.

Another important (and closely related) implication of the moralized-and-systematic view is that the harm principle is *flexible* in the sense that a wide range of first-order views are compatible with the harm principle. Traditionally, the harm principle has been associated with a recognizably Millian brand of liberalism—one that is highly opposed to paternalistic laws and policies. But if the harm principle is merely shorthand for some prior moral or political theory, then there seems to be no reason for it to be the exclusive preserve of a certain sort of Millian liberal. Instead, a wide range of first-order views can be married to the harm principle, such as Kantianism, Rawlsianism, utilitarianism and libertarianism. From a Rawlsian perspective, for instance, Jonathan Quong explains that "to the extent some version of the harm principle is a valid principle of liberal justice, I believe it is fundamentally justified by appeal to the moral status of persons, that is, the harm principle is generally a sound principle of political morality because it ensures that citizens are not treated in a certain type of paternalistic fashion, one which demeans their moral status as free and equal persons".[17]

This flexibility is illustrated well by Raz's reconciliation of the harm principle with the sort of perfectionist political theory that is often assumed to be inimical to the harm principle. Indeed, one can see Raz's argument as a kind of formula or template for reconciling the harm principle with some or another systematic normative theory.[18] Recall Raz's summary of his argument:

---

[16] Brink, D., *Mill's Progressive Principles* (Oxford: Oxford University Press, 2013), p. 190.
[17] Quong, J., *Liberalism Without Perfection*, pp. 55-6.
[18] Feinberg similarly explains that legislators employing the harm principle "will have to use some supplementary criteria because otherwise the harm principle is a *mere empty receptacle*, awaiting the provision of normative content before it can be of any use" (see *Harm to Others*, p. 245, emphasis added).

> The government has an obligation to create an environment providing individuals with an adequate range of [valuable] options and the opportunities to choose them. The duty arises out of people's interest in having a valuable autonomous life. Its violation will harm those it is meant to benefit. Therefore its fulfillment is consistent with the harm principle.[19]

This allows even defenders of moderate paternalism to lay claim to upholding the harm principle by filling out Raz's formula as follows:

> The government has an obligation to create an environment in which moderate paternalistic laws and policies (such as seatbelt mandates) protect individuals from their own occasional weaknesses of reason and will. The duty arises out of people's interest in enjoying health, wealth and other components of well-being. Its violation will harm those it is meant to benefit. Therefore its fulfilment is consistent with the harm principle.

Exactly *how* flexible is the harm principle according to the moralized-and-systematic view? One might wonder whether it allow defenders of the sort of theocratic absolutism evident in certain periods of medieval European history to lay claim to upholding the harm principle by filling out Raz's formula in the following way:

> The government has an obligation to create an environment in which inquisitions and other ecclesiastical tribunals combat heresy and the worship of false gods. This duty arises out of people's interest in enjoying eternal salvation and in avoiding eternal damnation. Its violation will harm those it is meant to benefit. Therefore its fulfilment is consistent with the harm principle.[20]

This case illustrates that some *constrains* need to be imposed on the sorts of first-order normative theories that can be wedded to the harm principle. Different constraints are possible and we need not adjudicate between them here. One constraint is *minimal plausibility*: only first-order normative theories that are minimally plausible can be wedded to the harm principle. One can then rule out the reconciliation of the harm principle with theocratic absolutism either by arguing that it is not minimally plausible to suppose that there is an interest in enjoying eternal salvation and in avoiding eternal damnation or by arguing that it is not minimally plausible to suppose that this interest gives rise to a governmental duty to create an environment in which inquisitions and other ecclesiastical tribunals combat heresy and the worship of false gods (or both). Another constraint is *fidelity to the underlying motivations of the harm principle*: only first-order normative theories that remain faithful to the underlying motivations

[19] Raz, J., *The Morality of Freedom*, pp. 417-8.
[20] I thank Tom Christiano for suggesting this case.

of the harm principle can be wedded to the principle.[21] (Of course, if it is the case that a first-order normative theory is minimally plausible only if it remains faithful to the underlying motivations of the harm principle, then this fidelity-to-the-underlying-motivations constraint turns out to be entailed by the minimal-plausibility constraint.) And while there may of course be different interpretations of what the underlying motivations of the harm principle are and of what it takes to remain faithful to them, it is difficult to see how a theocratic-absolutist normative theory could possibly be regarded as faithful to these motivations.

Regardless of which constraint or set of constraints we select, the key point remains that the moralized-and-systematic view does not allow one to "bootstrap" *any old* normative theory into conformity with the harm principle merely by asserting the existence of some governmental duty or merely by asserting the existence some interest out of which that governmental duty is said to arise.[22] Reconciling a first-order normative theory with the harm principle calls for detailed philosophical argumentation, not mere stipulation. In this way, the harm principle is very flexible but not so flexible that it becomes completely *empty* or *vacuous*: there remain certain constraints on the sorts of first-order normative theories that can be wedded to the harm principle.

3.2 Against Alternative Views

Having set out in more depth my favoured view of the role of the harm principle, let me argue against the three alternative views found in the other quadrants of the two-by-two grid—namely, those that view the harm principle as moralized and unsystematic, as non-moralized and systematic, and as non-moralized and unsystematic.

3.2.1 The Harm Principle as Moralized and Unsystematic

On the moralized-and-unsystematic view, the definition of harm depends on other moral and political commitments (hence the *moralized*) but these commitments consist in a collection of piecemeal principles or maxims and do not take the form of any systematic normative theory such Kantianism, Rawlsianism, libertarianism or perfectionism (hence the *unsystematic*). Like the moralized-and-systematic view, this view also implies that the harm principle is redundant and flexible.

---

[21] In this vein Feinberg says in an important (but brief and somewhat cryptic) aside that an acceptable moralized definition of harm must have "some congeniality with the animating spirit of the harm principle insofar as it can be presumptively reconstructed" (see *Harm to Others*, p. 245).

[22] For a discussion of "bootstrapping" views into conformity with the harm principle, see Kramer, M., "Looking Back and Looking Ahead: Replies to the Contributors", in M. McBride and V. Kurki (eds), *Without Trimmings: The Legal, Moral, and Political Philosophy of Matthew Kramer* (Oxford: Oxford University Press, 2022), pp. 502-3.

A good illustration of the moralized-and-unsystematic is found in Feinberg's important and influential defence of the harm principle.[23] Feinberg holds that the definition of harm must refer to other moral and political commitments. Despite his slogan that a harm is a "setback to interests", Feinberg recognizes that harm (at least for the purposes of the harm principle) cannot be defined in a non-moralized way because some setbacks to interests do not violate rights, as in the case of a setback to interest that is consented to or that is incurred in legitimate competition. So Feinberg ends up settling on the official position that "a harm is a *wrongfully* set-back interest".[24] As he elsewhere puts it, "we have taken harm to be a state of set-back interests which is the product of the wrongful (right-violating) conduct of another party".[25]

However, Feinberg nowhere provides a systematic moral theory or conception of justice that could help us to determine in any given case whether a setback to interests is "wrongful" (or "right-violating") and therefore harmful. In fairness, he is aware of this omission. Describing his work as "applied moral philosophy", he makes the following methodological remark at the beginning of his work:

> Technical philosophers may find the approach in these volumes skewed…They will find no semblance of a complete moral system, no reduction of moral derivatives to moral primitives, no grounding of ultimate principles in self-evident truths, or in "the nature of man", the commandments of God, or the dialectic of history. It would be folly to speculate whether the moral theory implicit in this work is utilitarian, Kantian, Rawlsian, or whatever. I appeal at various places, quite unselfconsciously, to all the kinds of reasons normally produced in practical discourse, from efficiency and utility to fairness, coherence, and human rights. But I make no effort to derive some of these reasons from the others, or to rank them in terms of their degree of basicness. My omission is not due to any principled objections to "deep structure" theories (although I must confess to some skeptical inclinations). I do not believe that such an approach is precluded, but only that it is unnecessary. Progress on the penultimate questions need not wait for solutions to the ultimate ones.[26]

Instead, Feinberg prefers to rely on a collection of fifteen "mediating maxims" to give content to the idea of a wrongful setback to interests.[27] These include the maxim that setbacks to "patently wicked or morbid interests" do not count as harms; the maxim that consensual harms do not count as harms (also known as the *volenti non fit injuria* maxim); that maxim that "below a certain threshold" trivial harms do not count as harms (also known as the *de minimis non curat lex* maxim); the maxim that "mere transitory disappointments, minor physical and mental 'hurts', and a miscellany of disliked states of mind, including various forms of offendedness,

---

[23] See Feinberg, J., *Harm to Others*.
[24] Ibid., p. 105 (emphasis in original).
[25] Ibid., p. 186.
[26] Ibid., pp. 4, 17-8.
[27] Ibid., pp. 214-17, 243-45.

anxiety, and boredom" do not count as harms; the maxim that the creation of a "danger" is harmful in proportion to the "gravity of a possible harm", the "probability of harm" and the "valuable[ness] of the dangerous conduct"; and the maxim that setbacks to the interests of "abnormally vulnerable person[s]" as a result of "normally harmless actions" are not harmful.[28]

I doubt that there is any knock-down argument against Feinberg's view in particular or the moralized-and-unsystematic view in general. There is—to repurpose Rawls's remarks about intuitionism—"nothing intrinsically irrational" about unsystematic views like Feinberg's.[29] In the end, such views may well be true. We "cannot take it for granted that there must be a complete derivation of our judgments" about justice and moral rights from a single source.[30] Still, the fact that Feinberg's account is unsystematic does, I think, count against it at least to some degree. In part, its unsystematic nature is problematic for reasons of simplicity and parsimony. We should (other things equal) prefer views that posit fewer brute facts. So we should (other things equal) prefer views that define harm in terms of a systematic normative theory over views like Feinberg's that define harm in terms of fifteen mediating maxims. This becomes particularly clear if we imagine variations on (or developments of) Feinberg's view on which the number of mediating maxims rises to the twenties or thirties. In part, too, the unsystematic nature of Feinberg's account is problematic for reasons of illumination and insight. Feinberg's view leaves us wondering: What morally or politically significant feature do the fifteen mediating maxims have in common? What is the deeper rationale for why harm should be understood in light of these specific mediating maxims but not other possible maxims? To the extent that these questions remain unanswered, Feinberg's view does not really illuminate the concept of harm. We finish our reflections with not much more understanding of, or insight into, the nature and significance of harm than we had going into them.

In short, while we should recognize the possibility that something like Feinberg's plurality of distinct and irreducible maxims is the best we can do—while we should recognize, in other words, that there may simply be no informative answer to the question of why harm is determined by reference to all and only a dozen or so maxims—we should hope for more. This sort of appeal to an unconnected heap of mediating maxims should be seen as an explanatory last resort. At the very least, then, we would do well to take seriously views that embed the harm principle within some more systematic normative theory.

3.2.2 The Harm Principle as Non-moralized and Systematic

On this view, harm is defined in terms of some systematic non-moralized account of harm. Moralized accounts of harm—such as that of Raz mentioned in Section 1—hold that an action is harmful just in case it violates positive or negative duties. By contrast, non-moralized accounts of harm do not define harm in terms of moral rights or duties. Non-moralized accounts

[28] Ibid., pp. 215-7.
[29] Rawls, J., *A Theory of Justice* (Cambridge: Harvard University Press, 1971), p. 39.
[30] Ibid.

of harm can come in two main kinds: comparative and non-comparative. Comparative accounts of harm hold that an action is harmful just in case it makes a person worse off relative to some baseline. There are different sorts of comparative accounts depending on what one takes the baseline to be. According to the temporal comparative account of harm, an action is harmful just in case it makes a person worse off after the action than she was prior to the action.[31] According to the counterfactual comparative account of harm, an action is harmful just in case it makes a person worse off than she would otherwise have been.[32] Non-comparative accounts of harm hold that an action is harmful just in case it causes a person to be in an intrinsically bad state, such as a state of pain or disease or death.[33] Hybrid accounts of harm combine comparative and non-comparative elements. For instance, a recent hybrid account holds that an action is harmful just in case it is harmful either in the temporal sense or in the counterfactual sense.[34]

The non-moralized-and-systematic view promises to give the harm principle a non-redundant role to play in explaining or justifying individual rights and duties. Consider again the right to freedom of conscience. A proponent of this view would justify this right on the grounds that a person's thoughts and beliefs are harmless—either in the sense that they do not make anyone else worse off relative to some baseline (as per the comparative account of harm) or in the sense that they do not cause anyone else to be in an intrinsically bad state (as per the non-comparative account of harm) or in some combination of these senses (as per the hybrid account of harm). Of course, the *expression* of certain thoughts and beliefs, such as racist or sexist beliefs, can be harmful to others in one of these comparative, non-comparative or hybrid senses, especially in their cumulative effects. So there is a debate to be had about what qualifications a proponent of this approach to the harm principle would need to place on the right to freedom of expression. But the point remains that defining harm in terms of some systematic non-moralized account promises to give the harm principle a role to play in explaining what we should or should not do.[35]

---

[31] See, e.g., Foddy, B., "In Defense of a Temporal Account of Harm and Benefit", *American Philosophical Quarterly* 51 (2014), pp. 155-65; Zhou, Y., "What it Means to Suffer Harm", *Jurisprudence* 13 (2022), pp. 26-51.

[32] See, e.g., Klocksiem, J., "A Defense of the Counterfactual Comparative Account of Harm", *American Philosophical Quarterly* 49 (2012), pp. 285-300; Feit, N., "Harming by Failing to Benefit", *Ethical Theory and Moral Practice* 22 (2019), pp. 809-23.

[33] See, e.g., Shiffrin, S., "Wrongful Life, Procreative Responsibility, and the Significance of Harm", *Legal Theory* 5 (1999), pp. 117-48; Harman, E., "Harming as Causing Harm" in M. Roberts and D. Wasserman (eds), *Harming Future Persons* (London: Springer, 2009).

[34] See Unruh, C., "A Hybrid Account of Harm", *Australasian Journal of Philosophy* (forthcoming); for criticism, see Carlson, E., Johansson, J. and Risberg, O., "Unruh's Hybrid Account of Harm", *Theoria* (forthcoming).

[35] One might object that, even on the non-moralized-and-systematic view, harm is still strictly speaking redundant. After all, if the answer to the question of why a person's thoughts and beliefs are harmless is that they do not make anyone else worse off relative to some baseline or that they do not cause anyone else to be in an intrinsically bad state, don't *these* claims do all the explanatory work? What does the fact of harmlessness add to these claims? Here, again, the fact that a person's thoughts and beliefs are harmless is not what *makes* it the case that we should respect the right to freedom of conscience but is instead merely a *by-product* of the reasons in virtue of which we should respect the right to freedom of conscience—it is just that the relevantly fundamental reasons here are about what makes others "worse off relative to some baseline" or what causes a others to be in "an intrinsically bad state", rather than being about the "moral theory" that specifies "what is due" to each person (in Raz's case) or

The main problem with this view, however, is that the leading non-moralized accounts of harm all have extremely counterintuitive implications. All such accounts—be they temporal comparative, counterfactual comparative, non-comparative or hybrid—both under-generate and over-generate harm.[36] While it is not possible to attempt to demonstrate this exhaustively here, consider by way of illustration the counterfactual comparative account of harm. This account over-generates harm: it counts as harmful what seems clearly harmless. This can be shown through cases involving failures to benefit.[37] Suppose that A buys a coffee to give to B as a random act of kindness. Tempted by the smell of the coffee, though, A decides instead to drink it himself. The counterfactual account entails that A harms B by drinking the coffee. After all, B is worse off than he would have been if A had given B the coffee rather than drinking it himself. Yet it seems clear that A does not harm B. A merely fails to bestow a benefit on B, and merely failing to bestow a benefit on someone surely does not amount to harming that person. Even more troublingly—insofar as the harm principle into which this account is to be plugged is a necessary rather than a sufficient condition for the legitimate exercise of state power—the counterfactual account of harm also under-generates harm: it counts as harmless what seems clearly harmful. This can be shown through cases involving pre-emption.[38] Suppose that C breaks one of V's legs, but that if C had not done this then D would have broken both of V's legs. The counterfactual account entails that C does not harm V. After all, C's action does not make V worse off than V would otherwise have been: if C had not broken V's leg, then both of V's legs would have been broken. Yet it seems clear that C harms V. The fact that V would have been harmed to an even greater extent were it not for C's action seems irrelevant to whether C's breaking of V's leg is harmful.

Recently, Folland has sought to defend something like the non-moralized-and-systematic view against this problem.[39] She is aware that the leading non-moralized accounts of harm appear to both under-generate and over-generate harm. But she responds by rejecting "the assumption that [the harm principle] is plausible only if there is a full-blown, problem-free (at least to a

---

the "mediating maxims" that determine "wrongful (right-violating) conduct" (in Feinberg's case). I am quite sympathetic to this objection. I am inclined to think, that is, that the only way for harm to play a genuinely non-redundant explanatory role is to adopt the sort of view set out in Section 3.2.3 whereby harm is defined in terms of pre-theoretical intuitions about harm. But I do not need to rely on this objection as the non-moralized-and-unsystematic view faces other important problems. I am thus happy to grant, at least for the sake of argument, that appealing to facts about harm is somehow explanatorily necessary on the non-moralized-and-unsystematic view in a way that it is not on the moralized views discussed in Sections 3.1 and 3.2.1.

[36] For criticisms of leading accounts of harm, see, e.g., Holtug, N., "The Harm Principle", *Ethical Theory and Moral Practice* 5 (2002), pp. 257-89; Bradley, B., "Doing Away with Harm", *Philosophy and Phenomenological Research* 85 (2012), pp. 390-412; Petersen, T., "Being Worse Off: But in Comparison with What? On the Baseline Problem of Harm and the Harm Principle", *Res Publica* 20 (2014), pp. 199-214.

[37] For a recent attempt by a counterfactual theorist to overcome this problem by invoking the doing/allowing distinction, see Purves, D., "Harming as Making Worse Off", *Philosophical Studies* 176 (2019), pp. 2629-2656. This argument is criticised in Johansson, J. and Risberg, O., "Harming and Failing to Benefit: A Reply to Purves", *Philosophical Studies* 177 (2020), pp. 1539-48.

[38] For a rejection of various recent attempts by counterfactual theorists of harm to overcome this problem, see Johansson, J. and Risberg, O., "The Preemption Problem", *Philosophical Studies* 176 (2019), pp. 351-65.

[39] See Folland, A., "The Harm Principle and the Nature of Harm", *Utilitas* 34 (2022), pp. 139-53.

high degree) theory of harm that we can refer to".[40] The main reason Folland gives for rejecting this assumption is that "accepting a strong, extensive, demand in this spirit risks leading to a more widespread skepticism".[41] After all, the harm principle "is only one of many normative principles that appeal to the concept of harm".[42] Consider the doctrine of doing and allowing, according to which doing harm is more seriously wrong than merely allowing harm (of the same magnitude) to occur. This principle makes explicit reference to the concept of harm. Yet proponents of this principle almost never spell out full-blown, problem-free accounts of harm. Folland notes, for instance, that in the course of a recent defence of the doctrine of doing and allowing Fiona Woollard says that "there are questions about how 'harm' should be defined" but that she "intend[s] to remain agnostic about most of these questions" and that "for [her] present purposes, it is enough if we can recognize paradigm cases of harm; precise analysis of the concept of harm can wait till later".[43] So if the lack of a full-blown, problem-free account of harm undermines the harm principle, it would seem also to undermine other harm-invoking principles such as the doctrine of doing and allowing and the doctrine of double effect.

However, the analogy with other harm-invoking principles is questionable. Harm is the central and distinctive feature of the harm principle. By contrast, harm is not central to the doctrine of doing and allowing. Although it is true that the doctrine of doing and allowing does appeal to the concept of harm, the central and distinctive feature of that principle—the load-bearing feature, so to speak—is not harm but the distinction between doing and allowing. The failure of proponents of the harm principle to give an adequate account of harm is thus not like the failure of proponents of the doctrine of doing and allowing to give an adequate account of harm; it is like the failure of proponents of the doctrine of doing and allowing to give an adequate account of the distinction between doing and allowing. And such failure would, I think, be a serious problem for proponents of the doctrine of doing and allowing. Indeed, it is interesting to note in this regard that Woollard—whom Folland quotes as saying that proponents of the doctrine of doing and allowing can "remain agnostic" about the nature of harm since "it is enough if we can recognize paradigm cases of harm"—does not remain agnostic about the nature of the doing/allowing distinction and does not rely merely on our ability to recognize paradigm cases of doing and allowing.[44] On the contrary, she spends over seventy pages developing a detailed, multi-clause analysis of the distinction between doing and allowing—one that involves a range of further act-theoretic distinctions between behavior that is "part of" a harmful sequence and behavior that is "relevant to" such a sequence, between "positive" facts and "negative" facts, and so on.[45]

---

[40] Ibid., p. 150.
[41] Ibid., p. 151.
[42] Ibid., p. 149.
[43] Woollard, F., *Doing and Allowing Harm* (Oxford: Oxford University Press, 2015), p. 18.
[44] Folland, A., "The Harm Principle and the Nature of Harm", p. 150; Woollard, F., *Doing and Allowing Harm*, p. 18.
[45] For the complete statement of her analysis of the doing/allowing distinction, see Woollard, F., *Doing and Allowing Harm*, pp. 93-4.

In short, then, proponents of the harm principle have not offered an adequate systematic non-moralized account of harm—be it temporal comparative, counterfactual comparative, non-comparative or hybrid. And the failure by a proponent of the harm principle to offer an adequate account of harm is not (contra Folland) like the failure by a proponent of the doctrine of doing and allowing to offer an adequate account of harm. Instead, it is like the failure by a proponent of the doctrine of doing and allowing to offer an adequate account of the distinction between doing and allowing. And the failure by a proponent of the doctrine of doing and allowing to do this is a problem. So the failure by a proponent of the harm principle to offer an adequate account of harm is also a problem.

3.2.3 The Harm Principle as Non-moralized and Unsystematic

On this view, harm is defined in terms of our pre-theoretic intuitive judgements about harm in its natural-language sense. A key aspect of this view is that, as far as possible, it relies on *pre-theoretical* intuitions about harm—that is, intuitions that are not strongly shaped by a systematic moral or political theory, by a collection of piecemeal principles and maxims, or by a non-moralized account of harm—for otherwise it begins to shade into one of the other views about the harm principle. The hope is that we have a sort of independent grip on the concept of harm—one that allows us to form intuitions about harm that hang free from, and are neutral with respect to, various competing views about morality and justice.

There are different ways of developing this non-moralized-and-unsystematic view. One way is to treat harm as a kind of primitive or unanalyzable notion. On this version of the second view—and to use a phrase familiar from American obscenity law—harm as the sort of thing about which the most we can say is "I know it when I see it". Another way to develop this view is to emphasize that while there are disagreements about what counts as harmful, there is a set of "core" or "paradigmatic" judgements about harmfulness and harmlessness about which there is unanimous or at least near-unanimous agreement.[46] For instance, almost everyone can be expected to share the core intuition that "broken bones and stolen purses" (in Feinberg's phrase) are harmful, even if they have very divergent intuitions about the harmfulness or harmlessness of other forms of conduct.[47]

This view gives the harm principle an important and non-redundant role to play in explaining or justifying individual rights and duties. Consider once again the example of the right to freedom of conscience. On the non-moralized-and-unsystematic view, this right is justified on the basis that a person's thoughts and beliefs do not harm anyone else, where this claim about harmlessness is in turn established by consulting our pre-theoretical intuitions about harm. Thoughts are surely harmless, if anything is. In this vein, Thomas Jefferson states: "The legitimate powers of government extend to such acts only as are injurious to others. But it does

[46] For remarks in this vein, see Folland, A., "The Harm Principle and the Nature of Harm", pp. 150-1.
[47] Feinberg, J., *Harm to Others*, p. 214.

me no injury for my neighbor to say there are twenty gods, or no god. It neither picks my pocket nor breaks my leg".[48] In other words, on this view the explanation of why we should respect the right to freedom of conscience makes direct and essential reference to the intuition that a person's thoughts and beliefs are harmless—an intuition that is meant to satisfy the "I-know-it-when-I-see-it" test or that is meant to belong to our "core" or "paradigmatic" judgements about harm.

The main problem with this approach, however, is that the set of core judgements about harm will almost certainly be too limited to have application in a wide range of circumstances. Perhaps there will be unanimous, or at least near-unanimous, agreement about the most obvious cases, such as Feinberg's example of the harmfulness of "broken bones and stoles purses" or Jefferson's example of the harmlessness of "say[ing] there are twenty gods"—and in this limited range of circumstances appealing to harm may well be able to play an important and non-redundant role in explaining what we should or should not do. But as soon as we move beyond these most obvious cases, disagreements about harm will be rife. Different people have very different intuitions about whether offensive speech is harmful, whether quality-of-life offenses like loitering and public drunkenness are harmful, whether one can harm by omission, whether refusal to perform an easy rescue is harmful, whether failure to contribute one's fair share to the common good is harmful, whether events after one's death can be harmful, whether the corruption of moral character is harmful, whether consensual harms are harmful, whether microaggressions are harmful, whether trespass on privately owned land is harmful, whether small contributions to aggregate harms are harmful, and so on.[49]

Without some theoretical account of harm, we are left with no criteria for arbitrating these disputes in a rational manner. Each person can report her pre-theoretical intuitive judgements about harmfulness and harmlessness, but there is no principled and mutually intelligible way to determine whose judgements about such matters are correct. In these situations—as Rawls puts it in the context of the impasses of intuitions faced by intuitionists—"the means of rational discussion have come to an end".[50] Indeed, commentators on the "I-know-it-when-I-see-it" test within American obscenity law similarly worry that it can lead to irresolvable impasses: "In effect, 'I know it when I see it' can be paraphrased and unpacked as: 'I know it when I see it, and someone else will know it when they see it, but what they see and what they know may or may not be what I see and what I know'".[51]

4. Generalizing the Argument

[48] Jefferson, T., *Notes on the State of Virginia* (London: Penguin, 1999), p. 165.
[49] For a sense of the breadth and depth of disagreement over judgements about harm, see Epstein, R., "The Harm Principle—And How It Grew", *University of Toronto Law Journal* 45 (1995), pp. 369-417; Harcourt, B., "The Collapse of the Harm Principle", *Journal of Criminal Law and Criminology* 90 (1999), pp. 109-94
[50] Rawls, J., *A Theory of Justice* (Cambridge: Harvard University Press, 1971), p. 41.
[51] Goldberg, W., "Two Nations, One Web: Comparative Legal Approaches to Pornographic Obscenity by the United States and the United Kingdom", *Boston University Law Review* 90 (2010), p. 2123.

Let us turn to the wider relevance of some of the lines of argument developed in this paper. I shall seek to generalize, in an admittedly somewhat sketchy and speculative way, the foregoing discussion by suggesting that two-by-two grid of possible explanatory roles of the harm principle set out in Section 2—and in particular the moralized-and-systematic view defended in Section 3—can be used to shed light on the explanatory role played by other important and influential principles within moral and political philosophy.

For the purposes of illustrating how this generalization might go, consider the public justification principle that is central to contemporary theories of political liberalism and public reason liberalism. Modern democratic societies are characterized by reasonable disagreement about a range of important matters. Such disagreement, says Rawls, "is not a mere historical condition that may soon pass away" but is instead "a permanent feature of the public culture of democracy" and "the inevitable long-run result of the powers of human reason at work within the background of enduring free institutions".[52] In light of this, many liberal theorists defend the public justification principle. This principle says that the justification for laws and institutions must be acceptable to all reasonable citizens. On this view, political issues should be settled by appeal to shared political values such as freedom and equality, and not by appeal to controversial moral or religious doctrines such as Christianity, Confucianism and Aristotelianism. However true those doctrines may be, they are subject to reasonable disagreement within contemporary societies and as such cannot form the basis of a just and stable political regime.

As with different proponents of the harm principle, different proponents of the public justification principle can, I suggest, be helpfully understood as occupying positions within a two-by-two grid. Where a given account of the public justification principle falls within this grid depends on how it answers two questions about the definition of "reasonableness":

1. *The Moralization Question*: Does the definition of reasonableness depend on other moral and political commitments?

2. *The Systematicity Question*: Is the definition of reasonableness systematic?

As before, these questions are distinct and so they give rise to four possible views about the explanatory role of the public justification principle. The first view treats the public justification principle as *moralized* and *systematic*. On this view, reasonableness is defined in terms of some systematic moral or political theory. This is one way of interpreting John Rawls's view and is more explicitly defended by Jonathan Quong, who explains that "to say that certain principles of justice could be endorsed by all reasonable people is to say that those principles can be validly constructed from a normative ideal of society as a fair system of social

[52] Rawls, J., *Political Liberalism*, pp. 4, 36.

cooperation between free and equal citizens".[53] "Reasonable citizens," he continues, "are a hypothetical constituency defined in terms of their acceptance of this ideal".[54]

The second view treats the public justification principle as *moralized* and *unsystematic*. On this view, reasonableness is defined in terms of a collection of various piecemeal principles or maxims. John Rawls's view can also be interpreted in this way. Leif Wenar, for instance, argues that Rawls's definition of the "reasonable citizen" contains "five attributes"—the possession of the two moral powers of a sense of justice and a conception of the good, a willingness to propose and abide by fair terms of cooperation, a recognition of the burdens of judgement, the possession of a reasonable moral psychology, and an acceptance of the five elements of a constructivist conception of objectivity—where these five attributes do not seem to be unified in any deeper or more systematic way.[55]

The third view treats the public justification principle as *non-moralized* and *systematic*. On this view, reasonableness is defined in terms of some systematic procedural idealization of actual citizens. There are various possible idealizations of this kind, such as Gerald Gaus's standard of engaging in a "respectable amount of good reasoning" or Charles Larmore's standard of "thinking and conversing in good faith".[56]

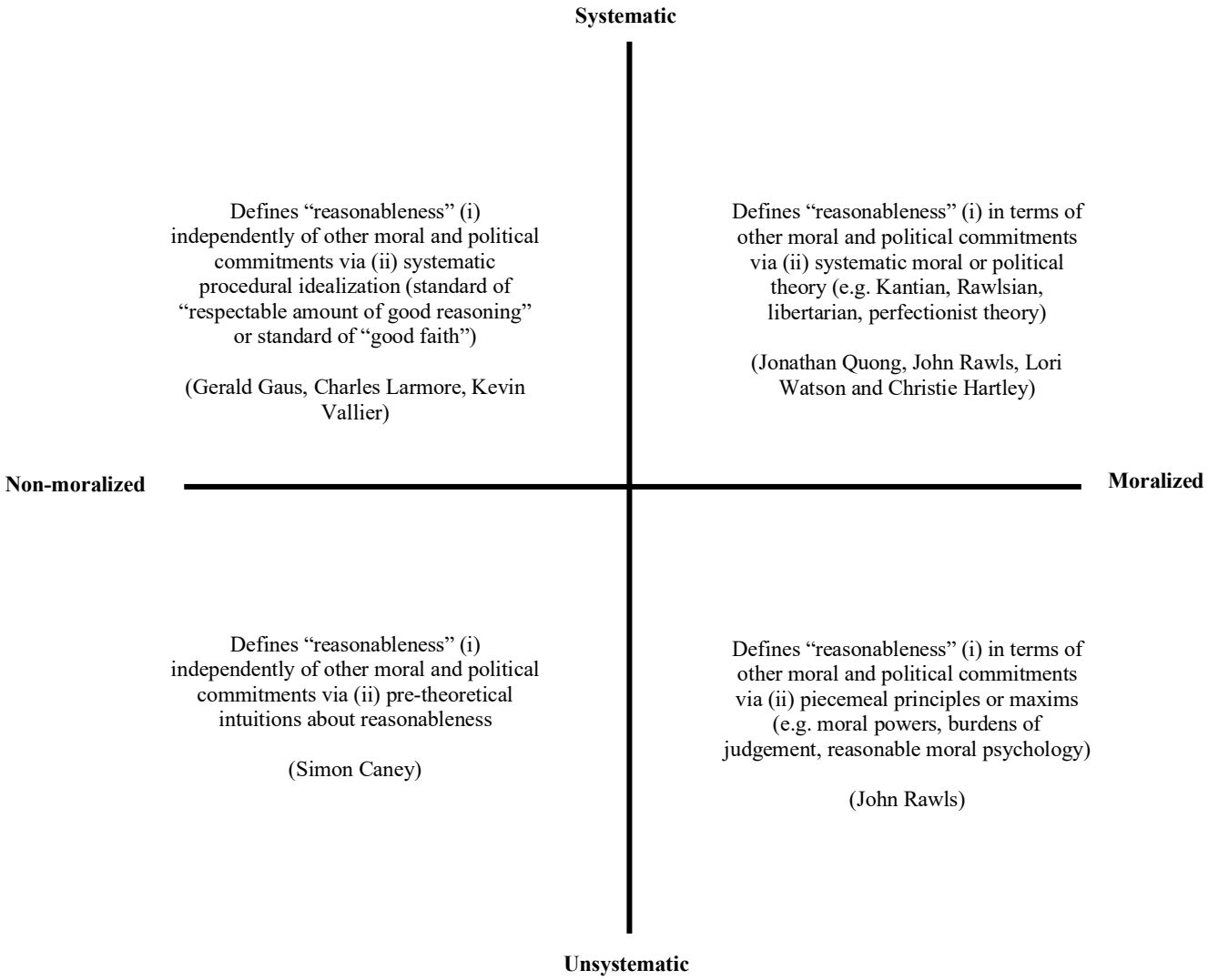
The fourth view treats the public justification principle as *non-moralized* and *unsystematic*. On this view, reasonableness is defined in terms of our pre-theoretical intuitive judgements about reasonableness in its natural-language sense. In the course of criticizing the public justification principle, Simon Caney appears to adopt this sort of view about the definition of reasonableness.[57] He considers a common argument against perfectionist laws and policies: (1) state action must be acceptable to all reasonable citizens (the public justification principle); (2) reasonable citizens disagree about what constitutes a good life (the fact of reasonable disagreement about the good life); so (3) state action must not be premised on claims about what constitutes a good life. He then argues against the second premise. After listing various implausible view about the good life—for instance, those that revolve around serious drug addiction and solvent abuse or those that revolve around extreme materialism and egoism—he asks rhetorically: "Can we plausibly argue that affirmation of these sets of beliefs about the good is reasonable?"[58] In this way, he says, "it seems implausible to claim that every judgement

---

[53] Quong, J., *Liberalism Without Perfection*, p. 144; see Rawls, J., *Political Liberalism*, pp. 39-40. This is also the view in Watson, L. and Hartley, C., *Equal Citizenship and Public Reason: A Feminist Political Liberalism* (Oxford: Oxford University Press, 2018).
[54] Ibid.
[55] Wenar, L., "Political Liberalism: An Internal Critique", *Ethics* 106 (1995), p. 37.
[56] See Gaus, G., *The Order of Public Reason: A Theory of Freedom and Morality in a Diverse and Bounded World* (New York: Cambridge University Press, 2011), p. 250; Larmore, C., "The Moral Basis of Political Liberalism", *The Journal of Philosophy* 96 (1999), p. 600. Gaus's standard of engaging in a respectable amount of good reasoning is endorsed by Kevin Vallier in *Must Politics Be War? Restoring Our Trust in the Open Society* (Oxford: Oxford University Press, 2019), p. 71 *et passim*.
[57] See Caney, S., "Impartiality and Liberal Neutrality", *Utilitas* 8 (1996), pp. 273-93.
[58] Ibid., p. 278.

about the good is subject to reasonable disagreement".[59] Importantly for our purposes, Caney's argument for the reasonableness or unreasonableness of certain judgements does not appeal to some moralized definition of reasonableness or to some non-moralized definition that incorporates a systematic procedural idealization; instead, it seems to appeal directly to pre-theoretical intuitions about reasonableness in its natural-language sense.

Once again, then, we can construct the following sort of grid of possible explanatory roles for the public justification principle:



Though I cannot show it here, I believe that that the explanatory role of the public justification principle is again most plausibly understood along the lines of the moralized-and-systematic view. As in the case of the harm principle, this view implies that the public justification principle is redundant (or at least very close to redundant) in the sense that it does not play any

[59] Ibid., p. 277.

role (or at least it plays only a very minimal role) in explaining what the state should or should not do.[60] It also implies that the public justification principle is highly flexible in the sense that it is compatible with a much wider range of first-order views than is typically supposed. Indeed, in other work I have argued for a reconciliation of perfectionism and the public justification principle through the idea of "perfectionist public reason". On this view, reasonableness is defined in terms of acceptance of the liberal values of freedom, equality and fairness as well as the perfectionist values of moral, intellectual and artistic excellence—where this definition of reasonableness is in turn grounded in a systematic perfectionist ideal of society as a fair striving for human flourishing between free and equal citizens.[61]

In short, then, it appears that what goes for the harm principle can be generalized, mutatis mutandis, to the public justification principle: different proponents of the public justification principle can (like different proponents of the harm principle) be helpfully organized within a two-by-two grid according to whether the definition of "reasonableness" depends on other moral and political commitments and is systematic, and the most plausible quadrant of this grid holds that the public justification principle does not itself explain the legitimacy or illegitimacy of particular institutions and laws but is instead shorthand for a systematic normative theory.

Of course, even if this grid in general, and the favoured quadrant in particular, can be generalized from the harm principle to the public justification principle, one might wonder how widely these ideas can be generalized. Can they be generalized to all, most or at least many other important and influential principles within moral and political philosophy, such as principles prohibiting exploitation, deception, manipulation and using other persons as a means?[62] But, assuming that these ideas can be widely generalized, it follows that many principles that are often thought to be more or less explanatorily fundamental—the harm principle, the public justification principle, the means principle, and so on—turn out to operate at a much less fundamental explanatory level. Instead, these principles—and the central concepts they invoke, such as "harm", "reasonableness", "using as a mere means", and so on—are shorthand for a systematic normative theory and are strictly speaking redundant in explaining what we should or should not do.

5. What is the Point of Mid-Level Normative Theory?

[60] As in Section 3.1, I add the qualifications in parentheses because I think some kind of fidelity to the underlying motivations of the public justification principle imposes a constraint on the sort of first-order views to which it can be wedded. For instance, an attempt to reconcile political Christianity and the public justification principle through the idea of "Christian public reason" (according to which all reasonable citizens must accept various Christian doctrines, such as the doctrine of the Trinity) cannot plausibly be regarded as faithful to these motivations. To this extent, the public justification principle is not entirely redundant. Still, the relevant constraint is very weak, and so the public justification principle is very close to being redundant in the sense that it plays only a very minimal role in explaining or justifying individual rights and duties.

[61] See Tahzib, C., *A Perfectionist Theory of Justice*, esp. ch. 7,

[62] I explore the generalization to various principles within deontological morality in Tahzib, C., "How Fundamental Are Deontological Principles?" (manuscript).

Let me close by commenting on an intriguing upshot of the foregoing argument. If the harm principle—as well as a range of other important principles, such as principles prohibiting exploitation, deception, manipulation and using other persons as a means—are just shorthand for a systematic normative theory, this poses an important challenge to the point of much mid-level normative theorizing. The foregoing argument, that is, suggests that we would do well to dispense with much mid-level normative theorizing and the central concepts that it cites such as "harm" and "reasonableness" in favour of a more direct inquiry into the relative plausibility of different systematic theories of morality and justice.

The challenge is not merely that the principles and concepts that figure within mid-level normative theorizing are *redundant* in the sense that if harm, reasonableness, exploitation, deception, manipulation and using as a means are defined in terms of a prior theory of morality and justice, then to say that something is harmful (or unreasonable, exploitative, and so on) cannot be what *makes* it the case that we should act in a certain way but is instead merely a *by-product* of the reasons in virtue of which we should act in a certain way. Redundancy is certainly *part* of the challenge to the point of mid-level normative theorizing. But it is not the full challenge because appealing to principles and concepts that are strictly speaking redundant in explaining how we should act might still be helpful in some way—for instance, by helping us to present this explanation in a maximally clear or focused way.

So the challenge is that mid-level normative theorizing is not merely redundant but in fact positively *unhelpful* in the sense that it invites confusion and misunderstanding and leads to the relevant moralizations being hidden from plain sight and remaining inadequately developed and defended.[63] After all, if in the course of moral reasoning and justification someone appeals to a mid-level normative concept such as harm or reasonableness or exploitation, it certainly does not look like she means thereby to appeal to some full-blown systematic theory such as Rawlsianism, Kantianism, libertarianism or perfectionism. Certainly, this is not what is meant by "harm", "reasonableness" and "exploitation" in their non-technical, natural-language sense. So isn't appeal to mid-level normative principles and concepts a recipe for confusion and misunderstanding and people talking past one another? Wouldn't it be clearer and more straightforward to just state our views? Why not derive conclusions directly from the relevant fundamental normative commitments without the confusing mid-level language? Don't frequent appeals to harm, reasonableness and so on distract from the really crucial questions about the structure, content and foundations of one's favoured systematic normative theory—questions that often remain hidden from plain sight and remain inadequately developed and defended? Is it any wonder that commentaries on the works of leading proponents of the harm principle, say, often involve speculative reconstructions of the systematic moral and political theories presupposed by, but not explicitly stated and defended in, these works?[64] What, in

[63] For similar thoughts in a related context, see Valentini, L., "Rethinking Moral Claim Rights", *Journal of Political Philosophy* (forthcoming).

[64] For utilitarian interpretations of Mill, see Gray, J., *Mill on Liberty: A Defence* (London: Routledge, 1996) and Crisp, R., *Routledge Guidebook to Mill on Utilitarianism*, ch. 8 and for a non-utilitarian interpretation of Mill, see

short, is the *value* or *point* of mid-level normative theorizing? If it is both redundant and unhelpfully misleading, why not remove mid-level principles and concepts from general philosophical circulation? Why not instead engage in a more direct inquiry into the relative plausibility of the prior normative theories presupposed by mid-level normative principles and concepts?

We have, then, what appears to be a general challenge to the value or point of mid-level normative theory. There have been some precursors to this challenge. But they have been highly *localized*. For instance, some commentators have pointed out that opting for a moralized definition of harm renders the machinery of the harm principle superfluous.[65] Similarly, some commentators have pointed out that opting for a moralized definition of reasonableness renders the machinery of the public justification principle superfluous.[66] But while the issues of redundancy and unhelpfulness have been recognized in these ways in the context of specific mid-level principles and concepts, their force as part of a fully *general* challenge to the value of mid-level normative theory has not, I think, hitherto been recognized. In an instructive discussion of the concept of harm, for instance, Ben Bradley remarks that since "nobody really needs to talk about harm, and doing so invites unnecessary confusion", we should "let harm go the way of phlogiston".[67] But in my view this remark can now be generalized: since nobody really needs to talk about *any* of the concepts within mid-level moral and political philosophy (harm, reasonableness, exploitation, deception, manipulation, using other persons as a means and so on), and since doing so invites unnecessary confusion, we should let all these mid-level normative concepts go the way of phlogiston and instead derive conclusions in a more direct way from general theories of morality and justice.

To meet this challenge, defenders of mid-level normative theorizing need to offer an explicit and well-developed account of the way in which mid-level principles and concepts play a useful or illuminating role within the methodology of reflective equilibrium. Despite the ubiquity of mid-level normative theorizing within contemporary moral and political philosophy, such accounts are surprisingly few and far between. In subsequent work, I hope to undertake an in-depth reconstruction and evaluation of various possible accounts of the value or point of mid-level normative theory.[68] Here are what I take to be the most promising such accounts: (1) that mid-level normative theory conveniently indicates points of convergence

---

Ten, C. L., *Mill on Liberty* (Oxford: Oxford University Press, 1980). For a Kantian reconstruction of the philosophical foundations of Feinberg's project, see Richards, D., "The Moral Foundations of Decriminalization", *Criminal Justice Ethics* 5 (1986), pp. 11-16.

[65] See, e.g., Holtug, N., "The Harm Principle"; Petersen, T., *Why Criminalize? New Perspectives on Normative Principles of Criminalization* (Cham: Springer, 2020), ch. 2; Stanton-Ife, J., "What is the Harm Principle For?", *Criminal Law and Philosophy* 10 (2016), pp. 329-353.

[66] See, e.g., See, e.g., Van Schoelandt, C., "Justification, Coercion, and the Place of Public Reason", *Philosophical Studies* 172 (2015), pp. 1037-41.

[67] Bradley, B., "Doing Away with Harm", p. 411.

[68] See "What is the Point of Mid-Level Normative Theory?" (manuscript).

between otherwise conflicting normative theories;[69] (2) that it is part and parcel of what Rawls calls "working from both ends";[70] (3) that it helps us to apply reflective equilibrium in a "modular" way rather than "all at once" by enabling comparison of a certain component of a certain general normative theory with the analogous component of a rival theory whilst holding fixed the rest of the relevant theories;[71] (4) that it helpfully gives us some interpretive "direction" or "steer" insofar as we try to render determinate the highly indeterminate commitments that constitute the basic building blocks of systematic normative theories in ways that are congenial to the spirit of mid-level concepts like harm, reasonableness, exploitation, deception, and so on;[72] and (5) that it provides an illuminating level of description for certain purposes, in much the way that, say, the level of description that cites biological concepts is illuminating for the purpose of understanding living organisms even if concepts within biology are strictly speaking reducible to concepts with chemistry or physics. To close, I shall briefly assess the first of these accounts—which I call the "points-of-convergence" account.

The points-of-convergence account holds that appealing to mid-level concepts can be helpful in drawing attention to points of convergence amongst otherwise conflicting systematic normative theories. Calling an action "harmful", for instance, can be a convenient way of saying that this action violates certain absolutely basic rights on which all normative theories (or at least all serious normative theories) converge, such as the right to life and the right to bodily integrity. Here the terminology of "harm", despite being strictly speaking explanatorily redundant, might nonetheless be helpful in indicating this point of convergence between disparate theories in a vivid, rhetorically forceful or memorable way. And the same might go for other mid-level concepts such as such as reasonableness, exploitation, deception, manipulation and using as a means.

However, there are two issues with the points-of-convergence account. The first is that appealing to mid-level concepts in this convergence-indicating way can still invite confusion and misunderstanding. After all, when people say that an action is "harmful", it is certainly not obvious that they mean thereby to say that the action in question violates a right on which all serious normative theories (including Kantianism, Rawlsianism, libertarianism, utilitarianism, perfectionism, and so on) converge. Again, this is certainly not what is meant by "harm" in its non-technical, natural-language sense. So any heuristic value in conveniently and vividly drawing attention to points of convergence between serious normative theories would need to be balanced against the heuristic disvalue of unclarity and confusion.

The second issue with the points-of-convergence account is that even if mid-level concepts can play a helpful role in these specific cases, these points of convergence are, I suspect, likely to

---

[69] For this idea in relation to the harm principle, see Krom, A., "The Harm Principle as a Mid-Level Principle? Three Problems from the Context of Infectious Disease Control", *Bioethics* 25 (2011), pp. 437-44.
[70] Rawls, J., *A Theory of Justice*, p. 20.
[71] I thank Jon Quong for this suggestion; for a related thought, see Scanlon, T., *What We Owe to Each Other* (Cambridge: Harvard University Press, 1998), p. 214.
[72] I thank Jon Quong for this suggestion.

be the exception rather than the norm. Perhaps, that is, all serious normative theories converge on a handful of absolutely basic rights which can be vividly and memorably represented by invoking mid-level concepts. But as soon as we move beyond these most obvious cases, serious normative theories can be expected to diverge and the appeal to mid-level concepts will go back to being both redundant and unhelpfully misleading. Certainly, in many of the sorts of debates in which mid-level principles and concepts have in practice been invoked—for instance, in debates over offensive speech and conduct and over paternalistic laws and policies—serious normative theories diverge in terms of which rights they recognize. Outside of the most obvious cases involving the violation of absolutely basic rights such as the right to life and the right to bodily integrity, then, it remains the case that when exploring crucial issues about morality and justice the question we should almost always be asking—and the question that gets obscured by frequent appeals to mid-level normative concepts—is about the relative overall plausibility of different first-order theories of morality and justice.

At the very least, then, the foregoing arguments serve to bring into view and make pressing a range of questions about the method and substance of moral and political philosophy that otherwise remain obscure. What is the point of invoking mid-level principles and concepts in moral reasoning and justification? How do they provide illumination or insight? If these principles and concepts presuppose some or another systematic normative theory, why not instead derive substantive conclusions directly from the basic commitments of the relevant normative theory? Wouldn't that be more straightforward and transparent? Why hide from plain sight the wider normative theory? Shouldn't we do away with mid-level normative theorizing and concepts it cites—harm, reasonableness, exploitation, deception, manipulation, using other persons as a means, and so on—in favour of a more direct and detailed analysis of the relative plausibility of different theories of morality and justice?