



**H.L.A. Hart's Punishment and Responsibility:
Introduction to the Second Edition (2008)**

by John Gardner
Professor of Jurisprudence
University of Oxford
<http://users.ox.ac.uk/~lawf0081>

This is an author eprint, which may not incorporate final edits.
The definitive version of the paper is published in

H.L.A. Hart, *Punishment and Responsibility: Essays in the
Philosophy of Law* (2nd ed., Oxford: Oxford University Press
2008) | [doi: 10.1093/acprof:oso/9780199534777.001.0001](https://doi.org/10.1093/acprof:oso/9780199534777.001.0001)

© 2008 the author

The typescript appears here with the consent of the publisher,
under the publisher's eprint policy, or by author's reserved rights.
Please do not quote from or cite to this eprint. Always use the
definitive version for quotation and citation.

H.L.A. Hart's *Punishment and Responsibility*

Introduction to the Second Edition

JOHN GARDNER

1. *Crime and punishment*

What good comes of criminal punishment? How does it help to make the world a better place? Criminal punishment, and more generally the criminal justice system that makes it possible, requires a huge investment of money, time, and energy. It has high costs and many casualties. If the system is to be justified, there must be compensating benefits. We had better know what they are and establish whether they are sufficient. That simple thought permeates *Punishment and Responsibility*, and animates the book's most widely-remembered essay, 'Prolegomenon to the Principles of Punishment' (chapter 1).¹

Put like this, you may say, Hart's thought is not only simple but uncontroversial. Any action or practice that has costs – and which does not? – needs to pay its way in countervailing benefits or else it cannot be defended. The tricky question is: What counts as a countervailing benefit? Hart thinks that a reduction in future wrongdoing² qualifies as a countervailing benefit of criminal punishment (p.8 and *passim*). The law's punitive

¹ Parenthetical chapter and page numbers refer to *Punishment and Responsibility* and are valid for both the first (1968) and the second (2008) edition.

² Hart uses other expressions ('offences', 'breaches of rules', etc.) to avoid what he regards as the excessively moralistic overtones of 'wrongdoing' and its cognates. But his preferred words have excessively legalistic overtones. See section 7 below for discussion of Hart's focus on punishment by law.

measures may contribute to this reduction in various ways: by deterrence, by public education, by incapacitation, by rehabilitation, and so on. But however it is brought about, the relevant benefit, thinks Hart, lies in the tally of future wrongs that will, thanks to the system, never be committed. The job of showing how much benefit there is in the system then shifts largely to empirical researchers who are in a position to tally the uncommitted wrongs that can credibly be attributed to the punitive measures in question. This explains Hart's emphasis on statistics concerning the effects of the death penalty in 'Murder and the Principles of Punishment' (chapter 3).³ The death penalty for murder, thinks Hart, can be justified only if (*inter alia*) thanks to its use a sufficient number of murders, or at any rate a sufficient number of sufficiently grave wrongs, go uncommitted.

But what about wrongs that have already been committed? Is it also a countervailing benefit of the criminal justice system that, thanks to its existence, wrongdoers whose wrongs are already *fait accompli* will suffer⁴ for their wrongs? Hart famously thinks not. For him the suffering of the punished wrongdoer, be he ever so guilty, is always a cost and never a benefit of the criminal justice system. Indeed the suffering of the punished wrongdoer is criminal punishment's most alarming cost, the one that creates the heaviest burden of justification for those who believe that the system should be maintained. On this point, as on several others, Hart sides with the classical utilitarianism of Jeremy Bentham.

³ Also in the appendix (pp. 268ff) which Hart added for the 1970 reprint.

⁴ Strictly speaking, either suffering or deprivation (e.g. of money or time) can be punitively inflicted. The two often coincide but need not do so. See J.D. Mabbott, 'Professor Flew on Punishment', *Philosophy* 30 (1955), pp. 256–265 at p. 257–8. To save words I will use 'suffering' to mean 'suffering or deprivation'. In doing so I do not mean to deny that one might regard the distinction between punitive suffering and punitive deprivation as morally significant. Mabbott, for example, congratulates modern Western societies on having largely replaced punitive suffering with punitive deprivation. (Alas the congratulations, if not naive, at least proved to be grossly premature.)

Contrary to a vulgar ‘retributive’ view that Hart plainly regards as barbaric there is nothing intrinsically appealing about any kind of suffering, even when it is punitively imposed, and so any genuine appeal that lies in a system to impose such suffering must be an instrumental appeal. It must be based on whatever advantageous consequences the suffering has (including the avoidance of later suffering), not on the suffering itself.

This broadly utilitarian approach to justifying punishment is sometimes called ‘forward-looking’ in contrast with the ‘backward-looking’ retributive view. Hart draws the contrast in these very terms in ‘Punishment and the Elimination of Responsibility’ (chapter 7, at p. 160). But we can already see that in some ways the terms are misleading. All justifications for punishment, indeed all justifications for anything, are forward-looking in the sense that they explain how the justified thing promises to make the world a better place, or at least to avoid its getting any worse. The special feature of the retributive view is not that it attempts to defy this axiom. If it were, the retributive view would be easy (or a lot easier than Hart finds it) to dismiss as irrational. The special feature of the retributive view, rather, is that it finds some intrinsic – not merely instrumental – value in a certain type of suffering, namely in suffering that is deserved.

The only ‘backward-looking’ feature of this view is a subsidiary feature. It comes of the distinctively retributive view about what makes suffering deserved. On the retributive view, suffering is deserved to the extent (and only to the extent?⁵) that

⁵ An optional feature of the view: there is no other way to deserve suffering. Perhaps people with bad characters also deserve to suffer, and the suffering they deserve is not exhausted by what they deserve *qua* guilty wrongdoers, let alone by what they deserve by way of punishment for their guilty wrongs. See Thomas Hurka, ‘Desert: Individualistic and Holistic’ in S. Olsaretti (ed), *Desert and Justice* (2003). This reminds us that the logic of desert does not single out punishment as the only unwelcome thing one might deserve. As well as a punishment, one might also deserve a bad reputation, a fall from

it is borne by the guilty in proportion to (and on the ground of:⁶) their guilt, where ‘guilt’ designates a relationship that a wrongdoer has or had to a wrong that she already committed. This suffering-of-the-guilty is a good to be pursued, even if all other suffering is an evil to be avoided.⁷ We can see here a sense

grace, a stern warning, etc. There is no distinctive retributivist answer to the question of when these various things might be deserved. The question to which retributivism gives an answer is the question of when one deserves to *suffer* and the retributive answer is ‘in proportion to one’s guilt as a wrongdoer’. One kind of strict retributivist is inclined to add ‘and only in proportion to one’s guilt as a wrongdoer’. For an excellent defence of the diversity of deserts, showing how far retributivism is from having a monopoly, see Fred Feldman, ‘Desert: Reconsideration of Some Received Wisdom’, *Mind* 104 (1995), pp. 63–77.

⁶ Another optional feature of the view: the suffering is deserved, and hence intrinsically good, only if it is deliberately imposed by someone on the ground of the guilty wrongdoer’s wrong – in other words, only if it is inflicted on the guilty *as a punishment*. This narrow retributivism, unlike its broader counterpart, ascribes no intrinsic value to the suffering that fortuitously befalls guilty wrongdoers in cases of so-called ‘poetic justice’, or the suffering that they ‘bring upon themselves’ without the intervention of a punisher. Since Hart’s quarrel with retributivism extends to these narrow and broad variants alike, and since the two have indistinguishable implications for the justification of punishment, the distinction between the two will not concern us here. When I speak of the good of ‘suffering-of-the-guilty’ I mean to cover both options: both the good of the guilty’s suffering (broad) and the good of the guilty’s suffering punitively (narrow). On this distinction see Lawrence H. Davis, ‘They Deserve to Suffer’, *Analysis* 32 (1972), pp. 136–140; also Robert Nozick, *Philosophical Explanations* (1981), pp. 369ff.

⁷ Here I am accounting for (although reformulating) the three features that Hart lists as characteristic of ‘retributive theory’ in ‘Postscript: Responsibility and Retribution’ (chapter 9) at p. 231. The three features as I interpret them: the suffering must fall upon a wrongdoer; the suffering must be proportionate to the wrongdoer’s guilt; the suffering (or its imposition) is intrinsically good if the previous two conditions are met. Some ‘deontological retributivists’ resist the ‘intrinsic good’ formulation, and indeed any formulation in terms of ‘good’. For excellent criticism of this resistance, see Michael Moore,

in which retributivists are consequentialists just like the classical utilitarians, and a sense in which they are not. They are consequentialists just like the classical utilitarians in that they identify a good to be pursued, perhaps even *ceteris paribus* maximised. In their case it is the good of suffering-of-the-guilty (which the classical utilitarians regard, like all other suffering, as an evil to be avoided). On the other hand they are non-consequentialists, and unlike the classical utilitarians, in that this good of theirs cannot be identified in an entirely action-independent way. To identify certain suffering as suffering-of-the-guilty, and hence as retributively good, one must always identify the already-committed wrong in respect of which the wrongdoer is guilty.⁸ Of course the wrong in question might still be a wrong on utilitarian (or more broadly instrumental) grounds. It need not be an intrinsic wrong. The question of what makes the action a wrong in the first place is beside the point. The point is that the suffering is redeemed – turned from bad to good – by its relationship to the doing of the wrong.

In spite of his misleading formulation in chapter 7, then, Hart's disagreement with the retributivists is not about whether a defence of criminal punishment must focus (in his words) on 'the future good we can do to society including the criminal' (p.159). Of course it must. The retributivist agrees. The disagreement is about what counts as a relevant future good. As he makes clear in chapter 1 (e.g. p.9), Hart finds it perfectly *intelligible* to cite suffering-of-the-guilty as a 'general justifying aim' for the criminal justice system as a whole. A society could indeed pursue, and even attempt to maximize, such suffering as an intrinsic good. Hart's objection to this society is not that we

'Justifying Retributivism', *Israel Law Review* 27 (1993), pp. 15–49. For more on deontology, see below (text at note 15).

⁸ Recall that according to narrow retributivists (see note 6 above) one must also identify a second human action to identify the relevant good, namely the action of the person who deliberately imposes the suffering (the punisher).

cannot make sense of its practice. His objection is that its practice is immoral. Pursuing suffering-of-the-guilty as an intrinsic good is immoral because considered intrinsically (i.e. aside from its consequences) suffering is always and only an evil.

2. *The relevance of guilt and innocence*

This attitude leaves Hart with the obvious problem, faced equally by Bentham and many intervening authors with similar moral sensibilities, of how the distinction between guilt and innocence can figure in sound principles for criminal punishment. What makes this distinction relevant to the use of punishment if it is not relevant to the intrinsic value or disvalue of the suffering that punishment by its nature involves?

Some writers have answered that punishment by its nature involves guilt as well as suffering. ‘Punishment of the innocent’, they say, is an oxymoron.⁹ In chapter 1 Hart labels this the ‘definitional stop’ argument and he gives it short shrift. True, by the nature of punishment, all punishment is for a wrong that, at the time of the punishment, has already been committed (p. 5).¹⁰ But one might be punished for a supposed wrong rather than an actual wrong (many punishments are meted out in the mistaken belief that a wrong was committed). And even where there is an actual wrong, punishment need not be inflicted upon the wrongdoer (history is littered with examples of vicarious and collective punishments). Nor, when the wrongdoer is punished, need the wrongdoer be, or even be thought to be, a guilty wrongdoer (the wrongdoer may be excused, and known to be excused, and yet still be punished). The punishment of innocents in all of these cases may be immoral but it is not oxymoronic.

⁹ The *locus classicus* of the argument (not cited by Hart) is A.M. Quinton ‘On Punishment’, *Analysis* 14 (1954), pp. 512–17.

¹⁰ Again, Hart says ‘offence’. See note 2 above.

Indeed – and this is the more important point emphasized by Hart (pp. 5–6) – claiming that the punishment of innocents is oxymoronic is evading, not answering, the question under investigation. For even if we were to accept that in logic there can be no punishment without guilt, we would still want to know: Why should we maintain a practice, the practice of punishment, that has this peculiar logic? Why not replace it with, for example, a practice of compulsory detention and treatment for the dangerous, never mind their guilt or innocent? It is an adequacy condition for any defence of the practice of punishment that it shows why the practice of punishment, inasmuch as it discriminates between the guilty and the innocent in meting out suffering, is better than an imaginable rival practice that metes out suffering irrespective of guilt and innocence, instead focusing exclusively on the instrumental value of the suffering in, say, preventing future wrongs. The ‘definitional stop’ argument does not meet this adequacy condition. It leaves those who want to defend punishment as a way of preventing future wrongs, but who want to protect the innocent in the process, powerless against those, such as Barbara Wootton, who say that punishment is irrational precisely because of this specious anxiety about guilt and innocence that it provokes. Hart devotes much of ‘Changing Conceptions of Responsibility’ (chapter 8) to fending off the Wootton challenge. As he shows, this challenge is a moral one that calls for a moral response, in the framing of which the ‘definitional stop’ is worse than useless.

Hart also makes light work of the solution favoured by Bentham himself, namely that the distinction between guilt and innocence directly affects the instrumental value of the suffering involved in punishment. Innocents, says Bentham, cannot be deterred. Hart’s reply is well-known and decisive. To conclude from the fact that innocents cannot be deterred by threats of punishment that the punishment of innocents cannot make a contribution to the deterrence of others is a ‘spectacular *non sequitur*’ on Bentham’s part (p.19, echoed on pp.43 and 77).

There is no rescuing Bentham from this objection. More promising and testing, however, is John Rawls' famous view that the distinction between guilt and innocence may have some *indirect* bearing on the instrumental value of punitive suffering.¹¹ To justify the punishment of a particular person for a particular wrong, says Rawls, we need to proceed in two stages. First we need to establish that the punishment has been meted out in accordance with the rules of the practice of punishment. Then we need to justify the practice of punishment. From the fact that the justification of the practice is entirely instrumental, Rawls points out, it does not follow that the rules of the practice require their users to reason instrumentally. Indeed the rules need not show any sign of their instrumentality on their faces, and, taken one at a time, may even defy instrumental justification. For perhaps it is only when working in combination with the other rules of the practice that these rules pay their way in good consequences. This being so, there is no reason to think that a retributive rule, or more generally a rule that has punishers distinguish the guilty from the innocent, could not be a rule of a wholly utilitarian practice of punishment, a practice that on the whole has more utility than any alternative, when we take account of the utility of the rules themselves and the utility of their combination and interaction. The utility of the rules may include, for example, the utility of reduced uncertainty and insecurity that comes of their use; the utility of their combination and interaction may include, for example, the utility of checks and balances, or more generally divided labour, as between different decision-makers.

Hart is not unsympathetic to this line of thought. He shares Rawls' view that the justification of punishment needs to proceed in (at least) two stages, and in particular that the question of how to justify the practice of punishment (Hart calls it the

¹¹ 'Two Concepts of Rules', *Philosophical Review* 64 (1955), pp 4–13.

question of 'general justifying aim') is distinct from the question of how punishments should be meted out within the practice (which Hart calls the question of 'distribution'). But Hart does not share the Rawlsian optimism that the rules of distribution, with their emphasis on the distinction between guilt and innocence, can adequately be defended by relying on their combined indirect contribution to a broadly utilitarian justification for the practice as a whole. Interestingly, Hart's objection to this view is not the standard one found repeatedly in the literature. The standard objection is that any broadly utilitarian defence of the rule against punishment of innocents leaves that rule too vulnerable to exigencies at the margins. One can always imagine extreme cases in which punishing the innocent would bring more benefit than following the rule would bring, even allowing for the value of the rule itself and the value of its combination with other rules. Many people are repelled from the Rawlsian line of thought by this possibility. But far from being repelled by it, Hart is drawn to it. We should not insist on upholding the rule against punishment of innocents come what may (p. 185). We should allow that the rule is sometimes overridden. The problem, says Hart, is not that the indirect utilitarian licenses such overrides. The problem is that when the indirect utilitarian licenses such overrides, she licenses them only as a sacrifice of utility (the utility of sticking to the rules of the practice) for the sake of greater utility (the utility of the action *in extremis*). There is nothing to regret in this sacrifice, any more than there is anything to regret in turning a smaller quantity of yoghurt into a larger quantity of qualitatively identical yoghurt. Whereas, says Hart, when we break the rule and punish the innocent, albeit justifiably, we should think of it as a sacrifice of something qualitatively different. The price we pay in such cases is not merely another quantity of utility that is swallowed up in the felicific calculus, but a quantity of some other value, a competing value that the Rawlsian line of thought fails to register but cannot in the end do without.

It seems to me that Hart misses Rawls' main point here. Rawls discusses punishment mainly in order to bring out a distinction between two ways of thinking about (justified) rules. On one account rules are mere summaries of what should be done according to the balance of underlying reasons, and are not themselves reasons for doing anything. But on the account that Rawls prefers and defends, rules make a rational difference. They are by their nature reasons for acting¹² and may sometimes be reasons for doing something other than what one should have done according to the balance of underlying reasons.¹³ If one adheres to the balance of underlying reasons in the face of a rule there is a rational price to pay for the adherence, in the form of a regrettable failure to conform to the rule. That remains true, one should add, even if all the reasons at stake in the conflict, including the rule itself, ultimately derive their force from the same value, e.g. the value of avoiding suffering. Regret is made rational, not by the existence of independent values that were not served, but by the presence of distinct reasons that were not conformed to.¹⁴ Even a single-value interpretation of practical thought therefore leaves logical space for rational regret once we factor in the rational appeal of pursuing that single value indirectly (through rules) as well as directly. It does not follow, of course, that Rawls succeeds in securing the guilt/innocence distinction its proper place in the practice of punishment. But it does follow that Hart does not have the argument he takes himself to have for rejecting the Rawlsian strategy.

Be that as it may, his rejection of the Rawlsian strategy leads Hart to abandon hope that the practice of criminal punishment

¹² This much Hart himself had argued in *The Concept of Law* (1961).

¹³ Later and more detailed defences of this Rawlsian view are to be found in Joseph Raz, *Practical Reason and Norms* (1975) and Frederick Schauer, *Playing by the Rules* (1991), among many other works.

¹⁴ See Bernard Williams, 'Ethical Consistency', *Proceedings of the Aristotelian Society (Supp. Vol.)* 29 (1965), pp. 103–124.

can be defended, in all of its aspects, as the instrument of just one single value. In particular, the central role of the guilt/innocence distinction in the distribution of criminal punishment cannot adequately be explained, Hart concludes, by pointing to (what he regards as) the practice's general justifying aim, viz. its part in the reduction of future wrongdoing. There must instead be some second and independent value that resides in reserving punishment for the guilty, a value that may compete with punishment's general justifying aim. Hart's technique for rooting out this value is to investigate, briefly in chapter 1 and then in much greater detail throughout the rest of the book, what qualifies as guilt for the purpose of criminal punishment.

In section 5 of this introduction I will discuss Hart's approach to this investigation. For now let me just report his main conclusions. The guilty, for the purpose of criminal punishment, are those who 'had, when they acted, the normal capacities, physical and mental, for doing what the law requires and abstaining from what it forbids, and a fair opportunity to exercise these capacities' (p. 152). By embodying this criterion of guilt, Hart says, the law makes it feasible for people to plan, or more generally to take steps, to avoid breaking the law, and thereby to avoid criminal punishments. Those whom the law punishes are the very same people to whom it already offered a fair way of steering clear of punishment. This transforms criminal punishment, says Hart, into a 'method of social control which maximizes individual freedom within the coercive framework of law' (p. 23). So here is the independent value we are looking for. Criminal punishment is oriented, in its general justifying aim, to the minimization of future wrongs, while simultaneously being oriented, in its main rule of distribution (or its main rule of justice, as Hart also frequently labels it), to the maximization of freedom. Of course, the two orientations may sometimes be at odds. Hart takes the view that the second yields a powerful but not absolute constraint on the pursuit of the first.

In the light of later work by Rawls, contemporary writers are apt to assume that rules of justice (unlike many other rules) are deontological. When people act justly they make the world a better place because their act is just. It is not that their act is just because it makes the world a better place.¹⁵ Hart rightly avoids this assumption. A rule of justice, he points out, is simply a rule ‘concerned with the adjustment of claims between a multiplicity of persons’ (p. 21). There is no reason why such a rule cannot be justified instrumentally, as Rawls himself had earlier shown.¹⁶ And sure enough, the case that Hart makes for the rule of justice that constrains punishment of the innocent is entirely instrumental. He maintains his Benthamite conviction that the entire practice of punishment (including any distributive rule of the practice) falls to be justified by its good consequences or not at all. He merely adds that there are (at least) two sets of good consequences of punishment to be considered, each a repository of an irreducibly different value. First there are the consequences of punishment for the incidence of wrongdoing, which explain (for the most part) why we should have such a practice at all, and then there are the consequences of punishment for the incidence of freedom, which explain (for the most part) why the practice should discriminate between guilty and innocent.

3. *The missing link*

You may object that freedom is the wrong kind of value for a consequentialist like Hart to espouse. That an action is freely

¹⁵ Rawls, *A Theory of Justice* (1971), at p. 30, restating a distinction first drawn in C. D. Broad, *Five Types of Ethical Theory* (1930), pp. 277–278.

¹⁶ Notice, however, that the fact that there is a rule can give the illusion of deontological orientation, because only someone’s doing something can qualify as conformity with a rule. Thus the value of conformity with the rule itself, such as it is, depends on someone’s doing something even though the value that underlies the rule and justifies its existence need not.

performed, where it lends value to an action at all, surely lends intrinsic rather than instrumental value to that action.¹⁷ So the value of the action no longer lies (only) in its consequences. True enough. So it appears that Hart is no more strictly a consequentialist than are his retributivist adversaries. He believes that certain human actions have intrinsic value or disvalue, value or disvalue quite apart from their consequences. It is no surprise at all to find this out. After all, Hart attributes intrinsic disvalue to suffering (even of the guilty) and suffering, as he notes, is among the constituents, not the consequences, of punishment (p. 4). It follows that Hart attributes intrinsic disvalue to the act of punishment. His refusal to attribute any redeeming (i.e. positive) intrinsic value to the same act does not come of any wider Benthamite view according to which human actions in general cannot have intrinsic value. Hart agrees with Bentham about suffering (it is in no way intrinsically good) and thus about punishment (it needs a wholly instrumental defence) but there is nothing in *Punishment and Responsibility* to suggest that, according to Hart, any other human action calls for a wholly instrumental defence (let alone a wholly instrumental assessment).

So Hart has no problem of consistency here. Nevertheless his instrumental defence of the punitive salience of guilt and innocence suffers from a serious weakness. The rule that it generates is, as we already noted, a *constraint* on the pursuit of punishment's general justifying aim. The pursuit of maximal freedom yields a rule against punishing the innocent,¹⁸ not a rule

¹⁷ Like J.S. Mill, Hart sometimes pays lip-service to the classical utilitarian instrumental defence of freedom (e.g. pp. 48–9 where he talks of the 'satisfactions' of choosing), but we know from his p. 12 objection to the Rawlsian strategy that this is not Hart's real position, any more than it is Mill's. On Mill's position, see C.L. Ten, 'Mill's Defence of Liberty' in J. Gray and G.W. Smith (eds), *J.S. Mill On Liberty: In Focus* (1991).

¹⁸ And possibly a rule against excessive punishment of the guilty, although (inasmuch as he supports it) Hart tends to support this 'proportionality' rule

in favour of punishing the guilty. It may of course be expressed, as I expressed it above, as a requirement to ‘reserve punishment for the guilty’ (or such like) but this formulation disguises the rule’s asymmetrical character. Equally misleading is Hart’s own characterization of the rule as allowing for ‘retribution in Distribution’ (p. 9).¹⁹ There is nothing even slightly retributive about Hart’s distributive rule, for under Hart’s rule the guilt of the guilty does not count in favour of punishing them; it merely eliminates an objection to punishing them. The only Hart-approved reason in favour of punishing the guilty (or anyone else) is the reason given by punishment’s general justifying aim, viz. that future wrongdoing is thereby reduced. Since one cannot properly act for a non-existent reason, it follows that securing a reduction in future wrongdoing is the only Hart-approved reason for which any punisher can properly act in punishing.

Unfortunately, this conclusion is inconsistent with the very nature of punishment as Hart himself explains it. As Hart himself explains it – and I think he explains it correctly – punishment must be *for* an (actual or supposed) wrong (p. 5), albeit the punishment need not be imposed on the (actual or supposed) wrongdoer. What does the italicized word ‘for’ mean here? It marks a supposed rational relationship between the wrong and the punishment. It means ‘by reason of’ or ‘on ground of’. So P punishes D only if the fact that D or another person (actually or supposedly) committed a wrong is among P’s reasons for making D suffer. To mount an adequate defence of punishment one must therefore mount an adequate defence of this reason: one must show how an already-committed wrong is a reason – at least a reason that could arise in some imaginable circumstances –

on different grounds. See chapter 7, esp. pp. 172–3. For further discussion see A.H. Goldman, ‘The Paradox of Punishment’, *Philosophy and Public Affairs* 9 (1979), pp. 42–58 at pp. 48f.

¹⁹ See John Cottingham, ‘Varieties of Retribution’, *Philosophical Quarterly* 29 (1979), pp. 238–46 at p. 241.

for P to make D suffer. Hart's defence of punishment does not meet this adequacy condition. It does not explain how there can truly be an element of 'retribution in Distribution'.

This shows that there was some vestigial insight in the far-fetched 'definitional stop' argument. Hart's brisk annihilation of that argument therefore comes back to haunt him. The definitional stoppers were indeed wrong to think that punishment must, by its nature, be imposed by reason of guilt. But they were right to think that punishment must, by its nature, be imposed by reason of actual or supposed wrongdoing (guilty or otherwise). Hart not only denies the existence of the first reason; he also, in the process, fails to acknowledge the existence of the second, and hence fails to account for what even he regards as a logically necessary feature of punishment. And while opponents of punishment like Wootton have nothing to fear from the discovery that this is a logically necessary feature of punishment, would-be defenders of punishment such as Hart cannot so easily brush it aside. Why? Because if one's would-be defence of punishment fails to defend a logically necessary aspect of punishment, then it is not a defence of punishment after all. One finds oneself in accidental alliance with those who say that the practice of punishment should be abolished and replaced with a practice that lacks this feature.

This is the awkward position in which Hart ultimately finds himself. He ends up advocating the replacement of punishment with a practice that resembles the practice of punishment in maintaining a distinction between the guilty and the innocent in the distribution of suffering, but that is quite unlike the practice of punishment in not treating the (actual or supposed) wrongdoing of the wrongdoer, even in cases of guilt, as a positive reason why the suffering should be inflicted. Hart fails to notice that there are here two distinct features of punishment that any adequate defence of the practice must defend. The first is dictated by morality and the second by logic. Having defended the first with his instrumental argument from freedom Hart takes

himself, wrongly, to have defended (or more likely obviated the need to defend) the second.

There is an element of wishful thinking, then, in Hart's talk of 'the decay of retributive ideas' (p. 180), and in his writing these off as ideas 'which we may very well discard' (p. 181). Hart fails to show these ideas to be either decadent or redundant. And inasmuch as he predicts their extinction in serious thinking about punishment, he could not be more wrong. Since the publication of *Punishment and Responsibility* a succession of serious and influential writers have argued afresh that the guilty wrongdoing of the guilty wrongdoer is part of (some say the whole of) the positive case for punishing him. By 1990 it was possible to claim, indeed, that 'the new retributivism has sounded the death-knell of traditional, consequentialist approaches to criminal justice.'²⁰ Leaving aside those who continued to speak up for the Rawlsian alternative (retributive rule, but no underlying retributive value), there have been two main strands of this 'new retributivism'. According to the rectificatory view, the punitive infliction of suffering upon the guilty is an annulment or confiscation of ill-gotten gains or ill-taken liberties.²¹ According to the expressive view, meanwhile, the punitive infliction of suffering upon the guilty is an emphatic way of expressing or communicating the judgment of guilt and thereby, in some versions, censuring the guilty wrongdoer and/or denouncing the guilty wrong.²²

²⁰ John Braithwaite and Philip Pettit, *Not Just Deserts: A Republican Theory of Criminal Justice* (1990), p. 209.

²¹ Some major defences: Herbert Morris, 'Persons and Punishment', *The Monist* 52 (1968), pp. 475–501; Jeffrie Murphy, 'Three Mistakes about Retributivism', *Analysis* 31 (1971), pp. 166–169; John Finnis, 'The Restoration of Retribution', *Analysis* 32 (1972), pp. 131–35; Michael Davis, 'How to Make the Punishment Fit the Crime', *Ethics* 93 (1983), pp. 726–52; George Sher, *Desert* (1987), chap. 5.

²² Some major defences: Andrew von Hirsch, *Doing Justice* (1976), chap. 8; R.A. Duff, *Trials and Punishments* (1986), chap. 9; Jean Hampton, 'An Expressive Theory of Retribution' in Cragg (ed), *Retribution and its Critics*

On both of these views, notice, punishment is to be borne by the guilty because of and in proportion to their guilt, where 'guilt' designates some relationship that a wrongdoer has or had to a wrong that she already committed. The fact of my guilt in wrongdoing is a positive reason why I should suffer – indeed part of punishment's general justifying aim – and does not merely eliminate an independent objection to my suffering based on my innocence, as it does in Hart's account. Moreover, on both views, the suffering of the guilty is sought not merely for the further good consequences of achieving it. More precisely, the infliction of suffering wholly or partly constitutes the relevant act of rectification or expression (as the case may be), and that act in turn is held to be intrinsically, not just instrumentally, valuable.²³ The retributivism here is no mere Rawlsian veneer. There can be little doubt about the authentically retributive credentials of these rectificatory and expressive views. And since the late 1960s these views have come to be influential not only in philosophy but also in penal policy. So Hart could scarcely have been more wrong in his assessment of retributivism's appeal and prospects.

(1992); John Tasioulas, 'Punishment and Repentance', *Philosophy* 81 (2006), pp. 279–322. Some think of Joel Feinberg as the *Urvater* of this strand of the new retributivism, but Hart is right to say that Feinberg's 'The Expressive Function of Punishment', *The Monist* 49 (1965), pp. 397–423 relies on expression of condemnation as a way of distinguishing a punishment from a mere 'penalty', not as a justification for punishing. Inasmuch as Feinberg offers any arguments in favour of punishment conceived expressively, they are instrumental in true Hartian style. See Hart's note on Feinberg (p. 263).

²³ Duff suggests that his communicative theory might be thought of as non-retributive because according to it punishment has a 'purpose beyond itself' in which it might fail, even though it serves this purpose non-instrumentally. See his 'Penal Communications: Recent Work in the Philosophy of Punishment', *Crime and Justice* 20 (1996), pp. 1–97 at p. 46. However this feature does not distinguish Duff's view from other credible retributive views. To say that, according to retributivism, the suffering of the guilty (or its infliction) is intrinsically valuable is not to say that it is non-derivatively valuable, or unconditionally valuable, or valuable in itself.

This is not the place to assess the philosophical success of either strand of the new retributivism, let alone to assess the quality of the new penal policies that attempted to reflect them. Our concern here is to assess the philosophical success of Hart's defence of punishment as represented in *Punishment and Responsibility*. But it must be said that the relationship between these various assessments has often been misunderstood. That Hart did not succeed in discrediting retribution as an aim for punishment is widely appreciated. But the explanation is often garbled. Many attribute Hart's failure to the fact that he daringly offered what is known as a 'mixed theory'²⁴ of punishment, relying on a conjunction of independent and partly conflicting considerations, reflecting an irreducible plurality of values. This is often thought to lend some kind of instability or incoherence to his ideas.²⁵ But Hart's defence of punishment does not suffer from any instability or incoherence. It suffers only from incompleteness. The real problem, to put it another way, is not that the Hartian defence of punishment is too mixed but that it is not mixed enough. To supply a complete defence of punishment Hart needs to add yet further considerations to those that he already marshals. In particular, he needs to explain how the fact that a wrong was committed can ever be a positive reason for inflicting suffering on someone. He needs to stir a more authentically retributive ingredient into his mix.

Where can the extra ingredient be found? My own sense is that the expressive strand of the new retributivism provides the

²⁴ Confusingly, some people like to say that Rawls also had a 'mixed theory' of punishment even though Hart's quarrel with Rawls is precisely that Rawls' account is a single-value one. See e.g. Nicola Lacey, *State Punishment: Political Principles and Community Values* (1988), pp. 46ff, Ted Honderich, *Punishment: The Supposed Justifications Revisited* (2006), pp. 165–6. For more detailed comparison of Hart and Rawls, see Michael Lessnoff, 'Two Justifications of Punishment', *Philosophical Quarterly* 21 (1971), pp. 141–148.

²⁵ See Lacey, *State Punishment*, above note 24, p. 49;

best place to start, although in spite of much work already done there is much work still to be done to deepen its foundations. As I said, however, this is not the place to assess the merits of either strand of the new retributivism. For present purposes we have a more pressing question. How much of Hart's original Benthamite project could survive the addition of a more authentically retributive ingredient, say along expressive lines? In my view, quite a lot. It is a long way from the realization that a retributive ingredient is *necessary* to make a complete defence of the practice of punishment to the conclusion that a retributive ingredient is also *sufficient* to make a complete defence of the practice of punishment. It is perfectly possible to accept the former view while continuing to regard the latter view, in Benthamite spirit, as barbaric. The former view, after all, leaves open the possibility that nobody should be punished, however guilty, unless their punishment also works to reduce future wrongdoing, or has some similarly important beneficial consequences. Indeed one may well hold that, while there can be intrinsic value in the infliction of suffering upon guilty wrongdoers – say as an expression or communication of their guilt – this is dwarfed by the intrinsic disvalue of the same act as an infliction of suffering *tout court*. This seems a natural thing to think as soon as one considers how many other ways there are to express or communicate the guilt of the guilty. One can usually, for example, express or communicate the guilt of the guilty by announcing it, or by reproaching or denouncing in strong language. If one is going to express or communicate the guilt of the guilty instead by visiting suffering upon them, one needs further and weighty considerations in favour of choosing this singularly brutal way of making one's point.²⁶ One had better be able to point to major instrumental benefits of the infliction of

²⁶ This case for cutting retributive considerations down to size is made by Hart himself in *Law, Liberty and Morality* (Oxford 1963), p. 66. See also see John Cottingham, 'Varieties of Retribution', above note 19, at p. 245.

suffering or else it is hard to escape the accusation that the brutality is gratuitous. Here we have a simple way of folding retributive considerations into a defence of punishment without sacrificing the central Bentham-Hart-Rawls intuition that punishment stands or falls mainly on the further beneficial consequences of the suffering involved (and of its infliction, and of the rules for its distribution, and so on).²⁷

Apart from a vague mistrust of ‘mixed theories’ in general, I suppose there may be special worries about this proposed mix. As we have seen, retributive considerations have a special primacy in the defence of punishment. There are many possible reasons for punishing people, but the fact that a wrong was (supposedly or actually) committed is the only reason for punishing people of which it is true that one is not punishing people unless one acts for this reason.²⁸ The others are optional but this reason (call it reason R) is essential. Doesn’t it follow that reason R must carry special weight – maybe not to the exclusion of all others, but at least overriding them – in the reasoning of punishers? And doesn’t it follow in turn that any adequate defence of punishment must defend reason R as a reason for punishment capable, at least, of overriding all others? No, none of this follows. That one is not a punisher unless one acts for reason R does not entail that one is not a punisher unless R is one’s overriding reason for punishing. It entails only that one is not a punisher except to the extent that one treats the other reasons for which one acts as reasons to act for reason R (i.e. as reasons for punishing rather

²⁷ For a similar suggestion, as well as a good literature review, see C.L. Ten, *Crime, Guilt, and Punishment* (1987), chap. 4.

²⁸ It is also true, as Hart points out, that one is not punishing D unless one acts for the reason that one’s act will make D suffer. But this, unlike the reason mentioned in the text, is not a reason for punishing D. Rather it is a reason for doing certain other things to D, such as locking him up, taking his money, or spreading rumours about him, *in order to* punish him.

than just as reasons for inflicting suffering).²⁹ This is consistent with regarding reason R as a reason of trifling weight in itself, a reason that would never be sufficient on its own to defend any punitive action. I am not convinced, for my own part, that reason R is quite *this* trifling. I tend to think that it has enough force to justify some minor punishments on its own, without instrumental reinforcement. But I agree with Hart that its force is often exaggerated. In most cases of punishment, and certainly most cases of punishment exacted by law, retributive considerations need to be heavily supported by others.

4. Intentions and side-effects

Why does punishment carry such a heavy burden of justification? Why does it need so many considerations to be marshalled in its defence? I have been emphasising the suffering of the punished as a cost that calls for heavy compensating benefits. But so far I have made little of the fact that punitive suffering is deliberately inflicted. It is a (i) considered and (ii) intended response to wrongdoing. There are two distinct logical necessities here. Those who spontaneously lash out in fury often intend to inflict suffering on a wrongdoer for his wrong, but because they do so unthinkingly, this is not a form of punishment.³⁰ Meanwhile

²⁹ This shows what is 'artificial' (p. 166) about the distinction drawn in the German *Strafgesetzbuch* between punishment (in proportion to and by reason of guilt) and 'other measures' (for the protection of society or reform of the offender). Hart suggests that this is a 'barren piece of conceptualism' but it is much worse than that. It is a confused and damaging distinction.

³⁰ It may be a form of retaliation. Retaliation may be but need not be a considered response. When retaliation is a considered response it differs from punishment in not being a response to (actual or supposed) wrongdoing but a response to (actual or supposed) aggression against the responder (or an ally). A business may well retaliate against a competitor who aggressively cuts prices but this is not usually a punishment (because nobody suggests that the price

those who seek restitution or reparation or apology from a wrongdoer for his wrong may well think carefully about the suffering that they will thereby inflict, but so long as they do not intend it – so long as the suffering is but a known side-effect of their intended actions – their actions too are non-punitive. The second of these features will concern us here.

In ‘Intention and Punishment’ (chapter 5) Hart embraces the Benthamite view that, all else being equal, the known side-effects of an action are morally indistinguishable from the action’s intended effects. At any rate he cannot see a case for drawing a moral distinction here that would be robust enough to warrant marking the distinction in the criminal law. Yet his own Benthamite reactions to the practice of punishment surely ought to have given him further pause for thought about this conclusion. Isn’t it part of what makes punishment so hard to justify, for Bentham and Hart alike, that the punisher by definition intends the suffering that she metes out? Isn’t it likewise part of what makes the retributive view so barbaric to both Bentham’s and Hart’s eyes that the retributivist is not appalled by this defining intention of the punisher, but welcomes it? Recall that Hart himself sets up the problem for would-be defenders of punishment as a problem, not about punishment’s effects, but about punishment’s aims, in particular its ‘general justifying aim’.³¹ His quarrel with retributivism is about whether the infliction of suffering for wrongdoing is itself a defensible general justifying aim for punishment. His objection is not, of course, that the infliction of suffering for wrongdoing is not a general aim of punishment. He agrees that it is. Indeed he regards it as a defining aim of punishment. The problem is that it is also, to Hart’s way of thinking, an intrinsically obnoxious aim, and so

cut is wrongful). I owe this point to Richard Brooks, ‘Threats and Punishment’, *Philosophy and Public Affairs* 17 (1988), pp. 235–239 at p. 236.

³¹ I have defended the view that justification always depends upon aims or intentions in my *Offences and Defences* (2007), chap. 5.

cannot in itself help to justify anything of which it is the defining aim. A further general aim, he insists, is needed to do the justificatory work.

Of course, Hart forgets that the further general aim needs to justify, not only the infliction of suffering, but also the defining aim of the punisher in inflicting it. This is the problem we discussed in section 3 above. The problem we are discussing now is different. It is that all Hart's repeated emphasis on punishment's aims – its intended effects – sits ill with the chapter 5 doctrine according to which the distinction between intended effects and predictable side-effects is irrelevant in determining the wrongness of an action, or more generally the moral position of a wrongdoer. Since this distinction affects the moral position of punishers why might it not similarly affect the moral position of those to be punished, and hence sometimes be reflected in distinctions between wrongs and (in law) between crimes?

Hart's critique of the moral salience of the distinction between intended effects known side-effects relies heavily on the supposed tendentiousness of the so-called 'doctrine of double-effect' found in Roman Catholic moral teachings. The type of case discussed by Hart, perennially discussed by advocates and critics of the doctrine, is one where the effects in question lie on the borderline between intended effects and side-effects. If one intends to crush a foetus's head (say, as part of an operation to save the mother's life), does one by that token intend the foetus's death or can the death be classified as a mere side-effect of the head-crushing? If one intends to dynamite a fat man out of a cave exit where he is stuck (say, in order to unblock the exit for five others trapped in rising water behind him) does one by that token intend to blow the fat man to pieces or is his being blown to pieces merely a side-effect of his being blown out of the

rock?³² There is no straight answer. The distinction between intended effects and side-effects (known or otherwise) breaks down in these cases. There is tendentiousness in the doctrine of double effect when it is claimed (as it often is) to avoid the breakdown of the distinction by resolving cases such as these cleanly.³³ This urge to make a clean cut at the border, with no classificatory indeterminacies, reflects what Hart nicely calls ‘a legalistic conception of morality’ (p. 125).

That there are indeterminate borderline cases, however, does nothing to suggest (*pace* Hart) that the distinction between intended effects and side-effects is not morally salient. Indeed the idea that a distinction cannot be morally salient just because of borderline indeterminacies is itself symptomatic of a rather legalistic conception of morality. A more natural view to take of the craniotomy and dynamite cases is that these cases are morally indeterminate (and hence raise special moral worries) because they are classificatorily indeterminate. This view presupposes, rather than casting doubt on, the moral salience of the distinction between intended effects and side-effects. One possible task left over for the law, on this view, is to make the borderline sharper for legal purposes (e.g. for the purposes of distinguishing murder from manslaughter) by using artificial criteria that would rightly be dismissed as tendentious outside the law. The doctrine of double effect, as typically elaborated, supplies some such artificial criteria.³⁴ Hart’s mistake is to think that, since these sharpening

³² This particular example, not mentioned by Hart, is usually credited to Philippa Foot, ‘The Problem of Abortion and the Doctrine of Double Effect’, *Oxford Review* 5 (1967), pp. 5–15. Foot used it to test Hart’s critique of the doctrine in chapter 5, which had first been aired in the previous issue of the *Oxford Review*. She says that the example, offering ‘light relief’, is ‘well-known to philosophers’ and figured in conversations she had with Hart.

³³ See Alison McIntyre, ‘Doing Away with Double Effect’, *Ethics* 111 (2001), pp. 219–255.

³⁴ Foot wisely sets these borderline cases aside before explaining what she regards as the doctrine’s appeal: ‘The Problem of Abortion and the Doctrine

criteria are artificial, so must be the supposed moral salience of the distinction between intended effects and side-effects that they are being used to sharpen. Not so.

5. *Mens rea and the rule of law*

Hart's sidelining, at least for criminal-law purposes, of the distinction between the intended effects and the side-effects of our actions is only one of several similar moves in *Punishment and Responsibility*. Elsewhere in chapter 5 Hart upbraids those writers on criminal law who devote their energies to distinguishing between effects foreseen as certain and those foreseen as probable, and between effects foreseen as probable and those merely foreseen as possible. Such discussions tend to be 'very barren', he says, since any of these states of mind is 'usually enough for criminal liability' and the distinctions among them are 'in most cases immaterial' (p. 117). In 'Negligence, *Mens Rea*, and Criminal Responsibility' (chapter 6) he challenges with similar gusto those who venerate the distinction between foreseen effects and effects that were unforeseen only because the agent was not paying as much attention as he could and should have paid. While this distinction is sometimes used to set the limits of criminal liability ('recklessness' is criminal, 'negligence' is not) widespread fixation with it among criminal-law scholars reflects 'unexamined assumptions as to what the mind is and why its "states" are relevant to responsibility' (p. 149). And in 'Acts of Will and Responsibility' (chapter 4), although Hart does not cast

of Double Effect', above note 32. Some subsequent writers in the thrall of a legalistic conception of morality have, however, mistaken Foot's excellent diagnosis of the problem in these cases (the 'closeness' of the borderline effect to the intended effect) for a failed attempt at a solution: see e.g. Jonathan Glover, *Causing Death and Saving Lives* (1977) p. 89; Jonathan Bennett, 'Morality and Consequences', in S. McMurrin (ed.), *The Tanner Lectures on Human Values II* (1981), pp. 107–8.

doubt on the importance for the criminal law of distinguishing the voluntary from the involuntary, he rejects attempts to draw the distinction by pointing to a supposedly distinctive mental state, such as a 'desire for muscular movement' or 'volition', that figures always and only in voluntary agency.

In all of these contexts, thinks Hart, the mistake is much the same. It lies, not in the belief that the mental states of offenders matter, but in the failure to see *why* they matter:

The reason why ... strict liability is odious, and appears as a sacrifice of a valued principle which we should make, if at all, only for some overriding social good, is not merely because it amounts, as it does, to punishing those who did not at the time of acting 'have in their minds' the elements of foresight or desire for muscular movement. These psychological elements are not *in themselves* crucial although they are important as aspects of responsibility. What is crucial is that those whom we punish should have had, when they acted, the normal capacities, physical and mental, for doing what the law requires and abstaining from what it forbids, and a fair opportunity to exercise these capacities (p. 152).

This last sentence, you will remember, captures Hart's proposed criterion for distinguishing the guilty from the innocent for the purpose of distributing punishment and hence for the purpose of attaching criminal liability (his 'criterion of guilt' for short.) And you will recall that his case for this criterion of guilt is based on the contribution that use of this criterion makes to individual freedom. How exactly are the two ideas connected? Hart's argument is most fully set out in 'Legal Responsibility and Excuses' (chapter 2) although it is recapitulated at several other points in the book. I will call it the 'rule of law argument' because it assimilates the criminal law doctrine *actus non facit reum nisi mens sit rea* to a group of other protections (e.g. against retroactive, secret, and vague laws) that are afforded by the ideal known as the rule of law. It is through this ideal that the mental element in crime is connected with individual freedom.

According to the ideal of the rule of law, the law must be such that those subject to it can reliably be guided by it, either to avoid violating it or to build the legal consequences of having violated it into their thinking about what future actions may be open to them. People must be able to find out what the law is and to factor it into their practical deliberations. The law must avoid taking people by surprise, ambushing them, putting them into conflict with its requirements in such a way as to defeat their expectations and frustrate their plans. These are all different ways of expressing much the same understanding of the ideal. It is easy to see how retroactive, secret, and vague laws, constantly changing laws, laws that are enforced in a partisan way, and so forth, fall foul of the rule of law so understood. They make it impossible for one to factor the legal position reliably into one's thinking about what to do. Hart's argument is that laws criminalizing and punishing the innocent (meaning those who are innocent according to his criterion of guilt) flout the rule of law in much the same way. In particular, if there are no relevant 'mental conditions' (p. 35) or 'subjective element[s]' (p. 152) built into the crime as defined by law, if the crime can be committed 'without *mens rea*' (p. 76), if there is no need to prove 'intention or something like it' (p. 116), then there is nothing in the law reliably to alert the person about to violate the law that he is in the situation to which the law applies, let alone that he is about to violate it. In that case no amount of clarity, openness, prospectivity (etc.) in the law will enable the subject of that law reliably to be guided by it (either to avoid violating it or to take account of the consequences of violation). So some element of *mens rea* is required in the name of fair warning, a fair warning which in turn is required in the name of the rule of law.

So much for the connection between *mens rea* and the rule of law. The connection between the rule of law and individual freedom is more troublesome. We already noted that, according to Hart, a system of criminal justice that avoids punishing the innocent (including those who lack *mens rea*) 'maximizes

individual freedom within the coercive framework of law' (p. 23). Does Hart mean that this system of criminal justice maximizes individual freedom *tout court*? He clearly can't mean this. For he points out that a system of criminal justice may avoid punishing the innocent, and yet otherwise be such that people living under it are 'repressed' (p. 47) – which is to say, far from free. In such a situation the criminal justice system only has its 'badness mitigated' (p. 47) by the fact that it does not punish the innocent. So on Hart's account not-punishing-the-innocent is clearly not the only thing that the criminal justice system can do to contribute to making people free, and does not suffice by itself to maximize people's freedom. This ties in with a 'Marxist' objection to his argument that Hart anticipates a few pages later, according to which the freedom he is parading is a 'merely formal freedom, not real freedom, and leaves one free to starve' (p. 51). Hart rightly objects to this specious way of formulating the objection.³⁵ But he does see the objection's force. The freedom conferred by the no-punishment-of-the-innocents rule as Hart interprets it and defends it (i.e. using his criterion of guilt) may be of vanishingly small consolation to some people, given other advantages, including other freedoms, that they lack.³⁶

³⁵ The objection is either (a) that the price one pays for freedom might be starvation (not 'freedom to starve', but starvation); or (b) that more freedom from coercion might not yield a wider range of opportunities (i.e. a wider freedom to act). Objection (a) concerns the conflict between freedom and other values, such as the avoidance of suffering. Objection (b) concerns the conflict between two dimensions of freedom itself (neither of which is more 'formal' than the other). The familiar formulation reported by Hart is ambiguous (and trades on the ambiguity) between (a) and (b). In objecting to this formulation Hart was no doubt influenced by Isaiah Berlin's 1958 lecture 'Two Concepts of Liberty' in his *Four Essays on Liberty* (1969).

³⁶ The objection imagined by Hart is brought to life in Alan Norrie's *Law Ideology, and Punishment: Retrieval and Critique of the Liberal Ideal of Criminal Justice* (1991), pp. 154ff.

The imagined 'Marxist' objection is in effect an objection to the whole ideal of the rule of law, and not just to the new work that Hart tries to do with it in the criminal law. It reflects an aspect of the ideal which is well explained by Raz:

[T]he rule of law is a negative virtue in two senses: conformity to it does not cause good except through avoiding evil and the evil which is avoided is evil which could only have been caused by the law itself.³⁷

That the law of our country meets the requirements of the rule of law does not in itself maximize our freedom, 'formal' or otherwise. It merely protects us against the extra unfreedom that the law, as an institutionalized system of rules with a putatively comprehensive and supreme authority and a coercive apparatus to back it up, is uniquely (or maybe just especially?) capable of heaping upon us. This gives us a new way to read Hart's original proposal according to which a system of criminal justice that avoids punishing the innocent 'maximizes individual freedom within the coercive framework of law.' What Hart means, one may suppose, is that such a system (we should add 'assuming it also meets the other requirements of the rule of law') maximizes our freedom as against the law itself, enables us to deal with the law on maximally free terms. Or, to put it more revealingly, such a legal system minimizes that extra unfreedom that we are vulnerable to only (especially?) because we are subject to the law, namely the unfreedom of unexpectedly finding ourselves in violation of the legal rules and unexpectedly bearing the legal consequences of that violation.

Alas, this reading leaves Hart with several new problems. First, there is now the problem of explaining why the guilt/innocence distinction (and hence Hart's criterion of guilt) is supposed to be relevant only, or especially, to laws that attract

³⁷ Raz, 'The Rule of Law and its Virtue', *Law Quarterly Review* 93 (1977), pp. 195–211 at p. 206.

punitive consequences, such as criminal laws. In some cases, the reparative or restitutionary consequences of a private lawsuit may inflict worse suffering on wrongdoers (albeit only as a side-effect) than would a criminal prosecution and punishment for the same wrong. So why does Hart's rule of distribution have special application to punishment? Why is it not a rule applicable to legal consequences more generally? (Or maybe it is?³⁸ In which case has Hart made progress with the specific problems that he set out to tackle in *Punishment and Responsibility*? Has he even engaged with the retributive instinct?) Second, there is now a revived problem of explaining why Hart's rule of distribution is such a powerful (albeit not absolute) constraint on the pursuit of punishment's general justifying aim. For surely a constraint that turns out to be a minimize of localized unfreedom has *ceteris paribus* less benefit to offer, and hence *ceteris paribus* less force to exert in conflict with competing considerations, than a constraint that serves (as Hart perhaps led us to expect that this one would) as a maxi miser of freedom considered more globally?

Of course these questions are not faced by Hart alone. They do not bear exclusively on his defence of *mens rea*, or more generally on his defence of the rule against punishing the innocent. They can be raised equally about almost all of the familiar requirements of the rule of law (such as the rules against retroactivity and secrecy) when these requirements are defended as instruments of freedom. These requirements are often thought to apply with particular force to the criminal law, and with less force to private law. And they are often thought to operate as powerful (albeit not absolute) constraints on the law's pursuit of otherwise acceptable aims. How can their admittedly limited and negative contribution to the pursuit of individual freedom suffice to explain any of this? True, the fact that these questions can be

³⁸ But see pp. 131–2 where Hart seems to insist that the rule of distribution for compensation differs from the rule of distribution for punishment.

asked about most of the requirements of the rule of law supports Hart's thought that the requirement of *mens rea* should be regarded as one of the requirements of the rule of law. But that does not bring us any closer to the answers we need.

The answers, it seems to me, are to be found in two places, lying quite far apart. Hart points towards the first, but in a slightly misleading way. He writes (in an attempt to explain the power of his advertised constraint on punishing the innocent) of the 'protection of the individual against the claims of society', of the law 'respecting the individual as such, or at least as a choosing being', an idea which is 'very central in the notion of justice' (p. 49). This is slightly misleading in two ways. First, the reference to 'society' makes us think first and foremost of public morality, perhaps of the special responsibilities of government. Second, the reference to a 'choosing being' returns us in an unhelpful way to further reflection on individual freedom; it makes us think that Hart is just reiterating his main point. But the moral issue that Hart is gesturing towards here is more general. It is essentially the same issue that confronts the potholers in the cave, with their fat friend blocking the exit as the waters rise. The issue for them (dynamite him out or not?) is not an issue of public morality and nor is it especially connected with anyone's freedom. It is a perfectly general issue in moral mathematics.³⁹ When the value in one thing (e.g. a life or an artwork or a landscape) is pitted against the like value in each of several other like things, must the case always be resolved in favour of the several and against the one? Or does it sometimes depend on how the value in each

³⁹ Nicely framed by John Taurek in 'Should the Numbers Count?', *Philosophy and Public Affairs* 6 (1977), pp. 293–316, which has spawned a huge literature. See notably Derek Parfit, 'Innumerate Ethics', *Philosophy and Public Affairs* 7 (1978), pp. 285–301; F.M. Kamm, 'Equal Treatment and Equal Chances', *Philosophy and Public Affairs* 14 (1985), pp. 177–194; T.M. Scanlon, *What We Owe to Each Other* (1998), pp. 221–49; Joseph Raz, 'The Force of Numbers' in Anthony O'Hear (ed), *Modern Moral Philosophy* (2004), pp. 245–264.

thing is engaged with, e.g. by act rather than by omission, by intended effect rather than by side-effect, as accomplice rather than as principal,⁴⁰ calculatedly rather than spontaneously, reluctantly rather than wholeheartedly, and so on? If the mode of engagement matters is that because sometimes extra value (or disvalue) lies in engaging with a certain value in one way rather than another? Or is it perhaps nothing to do with any extra value, instead being something about the very logic of value (e.g. that by its nature it is to be respected first and promoted only second) or the very logic of rational agency (e.g. that one's rational relationship with one's own actions is necessarily different from one's rational relationship with the actions of others)? Clearly this is not the place to attempt an answer to these vast and deep questions. It is, however, the place to remind ourselves once again that punishment is intended by the punisher to inflict suffering on the punished, and that this element of intention can be expected to fortify any rule against its infliction, however that rule may otherwise be justified. This (it seems to me) helps to explain the heightened force of Hart's distributive rule, and other requirements of the rule of law, as constraints on criminalization and criminal punishment (as compared with their limited role as constraints on reparation and restitution, and more generally on private law). This is probably not the only quirk of moral mathematics that bears on the 'protection of the [one] against the claims of [the many]' in the context of the criminal law but it is an important one. It shows us how little space there is to raise special doubts about the morality of the criminal justice system (as compared with the civil justice system, the taxation system, etc.) once the moral difference between intended effects and known side-effects is denied or suppressed.

The second answer is quite different. It brings us back, from general issues in moral mathematics, to the special predicament

⁴⁰ See my *Offences and Defences*, above note 31, chap. 3.

of the legal system and the state. The rule of law is not only an instrument of individual freedom. It is also an instrument of the law's own effectiveness.⁴¹ These two sources of value in the rule of law do not wax and wane together. Arguably, subject to the proviso entered by Hart's imaginary 'Marxist', the rule of law has its highest value as an instrument of individual freedom when it is observed by an (otherwise) repressive legal system. Here the freedom to stay out of the clutches of the law, or at least to be able to plan for one's encounters with it, may be particularly desirable (p. 47). By contrast, the rule of law has its highest value as an instrument of legal effectiveness when it is observed by a morally upstanding legal system.⁴² The more awful a legal system, the less effective we should want it to be. So perhaps the rule of law gains value as an instrument of freedom just as it loses value as an instrument of legal effectiveness. Perhaps this helps to explain the remarkable constancy of the ideal's appeal across many different times and places where legal systems exist.

But all of this brackets the question of how exactly the rule of law serves as an instrument of legal effectiveness. Isn't it arguable that the most effective legal systems (those most successful in securing their policy objectives) have been those operated as reigns of terror, reveling in arbitrariness, exploiting human weaknesses, and triggering conditioned responses? So isn't the rule of law more of an *obstacle* to legal effectiveness?

Yes and no. It depends on what one means by 'legal effectiveness'. Hart memorably and helpfully distinguishes the effectiveness of law as a guide from the effectiveness of law as a goad. 'We must cease,' he says 'to regard the law simply as a system of stimuli goading the individual' (p.44). He is not

⁴¹ Raz, 'The Rule of Law and its Virtue', above note 37, at p. 208.

⁴² Of course, even in a repressive legal system, the maintenance of the rule of law may have high value *to the repressors* as an instrument of the law's effectiveness. My point in the text concerns the non-relative value of the rule of law, i.e. its capacity to make the world a better place.

denying that the law can be an effective goad. It is simply that this is not the specifically legal way of being effective. The specifically legal way of being effective is by guiding people, by providing them with authoritative rules which are effective in the relevant way if and only if they are followed. True, there normally have to be back-up threats, threats of legal consequences in the event of rule-violation. But the threats are there to motivate the following of the rules among those who would not otherwise be so motivated. So even these contribute to the specifically legal way of being effective.

Why does it matter whether a legal system is effective in this specifically legal way? Think again about the general justifying aim that Hart attributes to criminal law and criminal punishment, namely that of reducing future wrongdoing. Here the law has an aim that makes reference to the law's own rules. The law succeeds in this self-referring aim (i.e. is effective in relation to it) if and only if, thanks to the criminal justice system, people do not violate the rules that make up the criminal law. (That is how Hart himself interprets the aim: p. 8). Of course it might sometimes be sheer accident that people do not violate the legal rules in question. The legal rules may sometimes coincide with moral or prudential rules and people may be disinclined to do what the law forbids for moral or prudential reasons without caring about the law. But to the extent that the law contributes to the non-violation of its own rules – to the extent that the non-violation is not an accident – that is achieved by people following the legal rules (whether because they approve of the rules or to avoid the associated punishments or to impress their friends or for any other reason).⁴³ The general justifying aim of criminal law and criminal punishment, then, is one that can only be achieved by making the law such that its rules can be followed –

⁴³ I include the parenthetical caveat, in Hartian spirit, to guard against the mistaken view that all legal-rule-followers are legal-rule-approvers. Hart drew attention to the difference in *The Concept of Law* (1961), chap. 5.

can be used as guides – by those who are subject to them. This aim requires the paradigmatically legal kind of effectiveness. Hence it requires observance of the requirements of the rule of law, including the distributional rule advocated by Hart.

So it turns out that Hart's distributional rule owes at least part of its force to the general justifying aim that Hart ascribes to criminal law and criminal punishment. The value of the rule against punishment of the innocent as an instrument of freedom is reinforced by its distinct value as an instrument of legal effectiveness, assuming that the legal rules thereby made effective are otherwise decent enough for their effectiveness to be something worth pursuing. Hart makes this distinct value seem rather indistinct by conflating the contrast between law and a goad and law as a guide with a different contrast between law as an 'economy of threats' (p. 40) and law as a 'choosing system' (p. 44). Thanks to the 'choosing system' proposal we are returned too quickly to an emphasis on the value of individual freedom, so that the distinct value of the rule of law as a contribution to legal effectiveness, as opposed to individual freedom, is obscured. In the process Hart elides an important part of the explanation of why the rule of law 'is virtually always of great moral value',⁴⁴ and hence yields powerful (but not absolute) constraints on the law's pursuit of its own objectives. He also unwittingly conceals and hence overlooks an important contribution that the defence of his general justifying aim for criminal punishment makes to the defence of his distributional constraint on criminal punishment. The two defences are thereby made to seem more independent of each other than they really are.

⁴⁴ Raz, 'The Rule of Law and its Virtue', above note 37, at p. 208.

6. Trouble with excuses

Fortified in these twin ways, the rule-of-law argument succeeds admirably as a defence of the criminal-law doctrine *actus non facit reum nisi mens sit rea*. In one respect, indeed, the rule of law argument sets a tougher standard for the criminal law than Hart is prepared to admit. Contrary to what Hart says, crimes of advertant *mens rea* ('recklessness') are more in keeping with the rule of law than are crimes of inadvertant *mens rea* ('negligence'). It is true, as Hart says in chapter 5, that in many cases people can properly be criticized for not 'exert[ing their] faculties'. This is a decisive answer to one who supposes that 'I just didn't think' is a good excuse (p. 134). It is also true, as Hart says on the same page, that the law can encourage people to exert their faculties by creating crimes of negligence. This strategy cannot be criticized, from the point of view of the rule of law, merely because it relies on the law's ability to goad (to trigger people's attentiveness through their knowledge of the law's attitude to inattentiveness) in support of the law's ability to guide (to get them to follow the rules). Nevertheless, where the other demands of the rule of law are met, the law's ability to guide a potential offender is more reliably secured by including, in the definition of each crime, at least one fact, material to the action's criminality, to which the defendant must advert if she is to be regarded as having committed that crime by that action. A fact that she could have become aware of by paying more attention is a possible substitute, but it is by no means a perfect substitute. It leaves more to chance in the warning that the law gives to the defendant of her impending violation. *Pace* Hart, the rule of law argument therefore militates in favour of including an element of advertent rather than inadvertent (or we might rather say: present rather than absent, active rather than omissive) *mens rea* in the definition of each crime. That 'I just didn't think' is not normally a good excuse for wrongdoing, and should not be treated as one by the criminal law, does not alter this fact.

The distinction I have just drawn – between ‘I just didn’t think’ as a good denial of *mens rea* and ‘I just didn’t think’ as a bad excuse – is missing from *Punishment and Responsibility*. Hart supposes that his rule-of-law argument, which serves so well as a defence of *mens rea* elements in criminal offences, can be used equally well to defend the place of excuses in the criminal law. Indeed he does not distinguish the two tasks. At the beginning of chapter 2, he says that the ‘mental conditions’ on which criminal liability often depends can also ‘be expressed in negative form as *excusing conditions*’ (p. 28, emphasis in original). He goes on to sweep the criminal law’s whole ‘system of attaching excusing conditions’ into the same equation (p. 47). Denying that one had the *mens rea* required to have committed a criminal wrong is, according to Hart, only presentationally different from offering an excuse as a defence to the charge that one committed that same wrong.⁴⁵ There is no difference of logic between the two, and – maybe because? – there is no difference of rationale. Hart’s proposed criterion of guilt covers, and his rule-of-law argument explains, both the pleas that criminal courts treat as denials of *mens rea* for particular crimes (denials of intention, knowledge, foresight, etc.) and those that are normally classified as excusatory (duress, provocation, excessive self-defence, etc.).

This equation involves Hart in an implausible stretching of the ideal of the rule of law. One lacks legal guidance in the sense relevant to his rule-of-law argument, Hart thinks, if one cannot choose (which presumably means: reasonably choose) to follow

⁴⁵ In ‘The Ascription of Responsibility and Rights’, *Proceedings of the Aristotelian Society* 49 (1948), pp.171–194, Hart had lent towards the opposite view. The paper was omitted from *Punishment and Responsibility*, because (says Hart) he was persuaded by his critics that it was wrong (p. v). But the two critics he mentions are critical of the omitted paper on a different point. They are P.T. Geach, ‘Ascriptivism’, *Philosophical Review* 69 (1960), p. 221–225 and G. Pitcher, ‘Hart on Action and Responsibility’, *Philosophical Review* 69 (1960), p. 226–235.

the law. But a more plausible view, to which I adhered in section 5 above, holds that one is enjoying legal guidance in the sense relevant to the rule of law so long as one can reliably factor the legal prohibition (the legal rule that makes one's action into a criminal wrong) and the legal consequences of its breach (the attendant legal rules about arrest, prosecution, punishment, etc.) into one's thinking about what to do. Those under duress need not miss out, in this sense, on the full guidance of the law. So long as their crime is one of *mens rea*, and the other requirements of the rule of law are met, they are as well-placed as the law can make them to factor the criminal prohibition and the potential legal consequences of its breach into their thinking about how to react to the threats that are being made against them by the person who is putting them under duress. That, indeed, is why the law bothers to regulate their case.

You may object: The law does not bother to regulate their case; on the contrary, it grants them an excuse. But an excuse, Hart rightly points out, is not an exception to the rule that makes the excused action a crime. It works on 'a different footing' from an exception (p. 14).⁴⁶ So here we see one way in which a case of duress is logically different from a case of absent *mens rea*. In the case of duress, unlike the case of absent *mens rea*, the defendant's action is still a criminal wrong but the defendant has a defence to that wrong. He is a wrongdoer but not a guilty one. So yes, the law regulates his case; his action is prohibited. And yes, the law

⁴⁶ Hart says, on the other hand, that a 'justification' defence, such as self-defence, is an exception to the rule that makes the 'justified' action a crime. The scare-quote marks around 'justification' are Hart's. He realises that the implication of his view is that these are not really justifications or defences at all, since there is no crime to justify or defend. This Hartian view of 'justifications' was subsequently defended at greater length by Paul Robinson in 'A Theory of Justification: Societal Harm as a Prerequisite of Criminal Liability', *UCLA Law Review* 23 (1975), pp. 266–292. For further discussion and citations see my *Offences and Defences*, above note 31, chaps 4 and 5.

expects the defendant to factor the legal prohibition into his thinking, as a consideration that may help to counteract the duressor's threats. It does not regard the defendant's situation as taking him outside the scope of legal guidance. Indeed if the defendant thinks of himself as outside the scope of legal guidance (i.e. as not covered by the prohibition) the law may take a dim view of it. Certainly if he adapts his behaviour to benefit from the excuse that he believes the law will grant him, the law is likely to react by denying him the excuse. That shows an interesting and important asymmetry in the ideal of the rule of law as applied to the criminal justice system. The rule of law requires that the law give to each of us scrupulous advance warning of our impending commission of a criminal wrong, so that we may be guided by the law in respect of that commission. But it does not require that we be given similarly scrupulous advance warning in respect of any excuses that may be available to us. For excuses are not there to guide us.⁴⁷ That is why Hart has to stretch the ideal of the rule of law so implausibly to bring excuses such as duress and provocation under it.

There is much more to be said about the nature and importance of excuses, and more generally about the nature and importance of defences, in law and in wider moral life. The literature on this subject has burgeoned since the late 1970s, and Hart's explanation of excuses in *Punishment and Responsibility* has often been doubted.⁴⁸ Many believe, as I do, that the difference

⁴⁷ See George Fletcher, 'The Nature of Justification' in S. Shute, J. Gardner and J. Horder (eds), *Action and Value in Criminal Law* (1993), pp. 175–186. Fletcher makes the argument in relation to justification defences, but what he says applies *a fortiori* to excuses.

⁴⁸ New subtlety was brought to the subject by George Fletcher's *Rethinking Criminal Law* (1977), pp. 552–579. Subsequent major contributions include Kent Greenawalt, 'The Perplexing Borders of Justification and Excuse', *Columbia Law Review*, 84 (1984), pp. 1897–1927; Michael Moore, 'Choice, Character, and Excuse', *Social Philosophy and Policy* 7 (1990), pp. 29–58; and

between an element of the wrong (such as a *mens rea* element) and a defence (such as an excuse) is not merely a presentational difference. However several of the arguments about excuses that have emerged in subsequent works are denied to Hart (or perhaps we should say: are self-denied by Hart). They proceed from consideration of excuses – their nature and importance – in moral life beyond the law. Hart shies away from arguments about moral life beyond the law. Of course, it is not that he wishes to avoid moral arguments. *Punishment and Responsibility* is a long series of moral arguments. But almost all of these are moral arguments of a relatively specialized kind. The problems of punishment and guilt are treated as answering mainly to a distinct set of moral considerations – both objectives and constraints – that bear mainly on the design of public institutions and legal rules. The resulting doctrines, such as the excusatory doctrines of the criminal law, are thereby cut loose from any similar or analogous doctrines that bear on the non-institutionalized parts of moral life, e.g. in the morality of personal relationships. What Hart offers is not a morality of guilt and punishment but only an *official* morality of guilt and punishment – a morality for public officials. Hart was one of a number of philosophers working in the 1960s, also including Rawls, who together revived political philosophy as a branch of moral philosophy (and helped to shape as well as to capture the distinctive liberalism of the 1960s) by asserting political philosophy's relative autonomy from the rest of moral philosophy. But one may wonder whether, in applying such an approach to punishment, Hart loses sight of some of the logic of his subject-matter. Doesn't the criminal justice system attempt, in its inevitably clumsy way, to institutionalize certain moral practices, including the practice of punishment with its

Jeremy Horder, *Excusing Crime* (2004), among many others. My own discussion occupies chaps 4–7 of my *Offences and Defences*, above note 31.

familiar relationships to wrongdoing and guilt, that already exist quite apart from the law and its institutions?

7. A political morality of punishment?

Another way to put the same question: Isn't the morality of state punishment dictated first by the morality of punishment in general, and only second (by way of modification) by the rule of law and similar specialized moral considerations? Hart already ventures a negative answer to this question in chapter 1, and thereby paves the way for his subsequent focus on the official morality of punishment. By its nature, says Hart, punishment is a paradigmatically legal practice, a reaction to 'an offence against legal rules' (p. 5). Punishment 'otherwise than by officials' is a 'sub-standard or secondary' case (p. 5). So anyone thinking about punishment, and hence about the criteria of guilt that make someone fit to be punished, should be predisposed to think about problems of official morality first, and to extend their thinking later to similar or analogous problems that arise elsewhere. I for one do not share Hart's conceptual intuition here, or even see where it gets its appeal.⁴⁹ Friends, colleagues, spouses, siblings and business partners regularly punish each other for actual or supposed wrongs that are not legal wrongs. They typically do so by withdrawing favours or co-operation, but there are many other possible ways, some of which are capable of involving the infliction of grave suffering. It is very common for one estranged spouse to punish the other, for example, by preventing him or her from spending time with his or her children, fully intending that this should be a terrible experience. I know of no reason to think that such punishment is 'sub-standard or secondary' as compared with, say, imprisonment by the courts. (Whether the

⁴⁹ For excellent discussion, see Leo Zaibert, *Punishment and Retribution* (2006), especially at pp. 22ff.

punisher must be engaged in a purported exercise of authority is another matter.⁵⁰ Perhaps so. But law, or officialdom, is equally not the paradigmatic exerciser of purported authority. Parents, deities, referees, arbitrators, religious leaders, teachers, and so on, are no less central cases of purported authority, irrespective of whether they owe or purport to owe their authority to law.)

State punishment is an important kind of punishment and no doubt it deserves special attention. But it deserves special attention mainly because it raises additional questions on top of those raised by the practice of punishment in general. Some of these questions are, of course, brilliantly raised and tackled by Hart. In particular, he succeeds from the very start of *Punishment and Responsibility* in focusing our attention on three important impacts that the legal institutionalization of punishment has on the morality of punishment so institutionalized.

First, the law tends to have the ability to influence people's behaviour on a much wider front than other punishers. Its influence is rarely restricted to the person punished, or to a small group of familiars. This means that the law can credibly have bolder objectives, in punishing, than the objectives that a friend or spouse or parent might credibly have. In particular, there is much more scope for the law to punish *pour décourager les autres* in significant numbers. This both strengthens the case for punishment and introduces extra temptations (such as the temptation to scapegoat), which in turn create a need for extra constraints to 'protect[] the individual' (p. 44).

Second, as we saw, the law exists to provide legal guidance and is required to provide scrupulous warnings to those who are about to break it. This is Hart's rule-of-law argument for *mens*

⁵⁰ For this weaker suggestion, see e.g. R.A. Duff, *Punishment, Communication, and Community* (2001), pp. xiv – iv. The requirement of purported authority may mark a difference between punishment and retaliation additional to the one identified in note 30 above. Or perhaps it is just a restatement or logical implication of that difference?

rea. This argument does not apply to non-institutional punishers. The non-institutional parts of practical thought are not regulated by the rule of law or any similar ideal, and the rules of non-institutionalized morality, in particular, are not tailored to restrict the element of surprise on the part of their violator that may accompany their violation. Many everyday moral rules, in particular, can be violated without any *mens rea* (although usually their violation can, on occasions, be excused). Those who act immorally and are punished for it by other non-institutional actors have no cause for complaint on this score, as they would have if they acted illegally and were punished by the law.

Third, the institutions of a legal system are bureaucratically organized and this affects the morality of punishment as administered by law. Different institutions have different roles in the punitive process. Some decide the range of punishments that would befit a certain class of wrongs, others select the alleged wrongdoers to be proposed for punishment (prosecuted), still others order the punishment of particular wrongdoers within that group, allocating punishments within the approved range, yet others supervise or carry out the ordered punishments, and so forth. This bureaucratic process sits well with Hart's proposal that there are various 'different questions' about punishment and that, in answer to these different questions, 'different principles (each of which may in a sense be called a "justification") are relevant at different points in any morally acceptable account of punishment' (p. 3). Different questions and different principles, Hart points out, are properly dominant at different stages of the bureaucratic process (p. 39). The legislature is most concerned with the general justifying aim of punishment, the judge with its distribution, and so on. This is also a theme, and a selling-point, of the Rawlsian defence of punishment.⁵¹ But it does not extend with the same force to non-institutional punishers, which are not

⁵¹ Rawls, 'Two Concepts of Rules', above note 11, at pp. 5–6.

bureaucratically organized (i.e. have no separation of powers). Many are just individual people punishing on their own.

Hart's 'different questions' approach has rightly been very influential in late-twentieth century thinking about punishment, especially the bureaucratic version that is criminal punishment. But Hart misses out a very important question from his initial list of 'different questions' about punishment that any student of punishment clearly needs to attend to. Who is to do the punishing? Hart does not ask the question because the answer – 'paradigmatically the law' – is already given, for him, by the very nature of punishment. This obscures from him the need to provide a defence of the way in which, and the limits within which, the criminal law takes over the business of punishing from other potential punishers. Many writers on criminal justice in the period since 1968, and indeed many policymakers, have worried about this 'theft' of conflicts by the law, conflicts which they think need to be somehow 'returned' to the victims of wrongs or their sympathizers or at any rate to non-bureaucratic players.⁵² Hart's contribution to the philosophy of punishment, otherwise probably the most important contribution of the twentieth century, has no answer to these worries except for his own curious version of the 'definitional stop': Punishment without law, he tells us, would only be punishment in a 'secondary or sub-standard' sense. But this is no answer to those who say: Fine, then paradigmatic punishment be damned. To answer these so-called 'abolitionists' – adapting from what Hart says about how to answer Wootton-style abolitionists – we need

⁵² The most influential discussion is Nils Christie, 'Conflicts as Property', *British Journal of Criminology* 27 (1977), pp. 1–15. See also Louk Hulsmann, 'The Abolitionist Case: Alternative Crime Policies', *Israel Law Review* 25 (1991), pp. 681–709.

to mount a moral argument defending the legitimacy of the law as punisher of first, or at the very least of last, resort.⁵³

How should we do this? Some extension of Hart's rule-of-law argument seems like a promising option. Isn't it part of the ideal of the rule of law that punishment should be monopolised by the law, or at any rate by the state? Yes and no. It is part of the ideal of the rule of law that punishment should be monopolised by the law in the sense that those who punish should be limited by law in when and how they punish, and that the law that limits them should also be capable of guiding them in doing so. But this same point applies *mutatis mutandis* to a wide range of exercises of power, and not only to punishments. Besides, it does not explain why punishments, or certain types of punishments such as those involving attacks on life and limb, or loss of liberty, or the appropriation of money, should be monopolised by the law in the stronger sense of being undertaken by the law itself – by legal officials – and not merely by ordinary people who are subject to the limits and guidance of the law in undertaking them. So much work is still needed to warrant the assumption made by Hart, and shared by many others, that 'Why punish?' is a question bearing first and foremost on the actions of public officials, rather than first and foremost on the actions of frustrated friends and despairing divorcees, albeit sometimes taken over by public officials.⁵⁴

⁵³ For constructive discussion of some possible problems with this undertaking see Thaddeus Metz, 'How to Reconcile Liberal Politics with Retributive Punishment', *Oxford Journal of Legal Studies* 27 (2007), pp. 1–23.

⁵⁴ My own contribution to justifying the role of public officials, especially in the criminal justice system, appears as chap. 11 of my *Offences and Defences*, above note 31.