1

Punishment

1. The Concept of Public Reason

Let us begin with a few definitions. As I use the term, coercive action can mean either of two things. It can mean the communication of a command backed by a believable threat of unacceptable harm if the addressee fails to comply with the command; or it can mean the use of physical force upon a person to carry out a threat or to pre-empt, confine, or direct its bodily movements whether or not there was an antecedent threat. When the commander is a public official authorized by the constitution to issue commands, the command is a law or order, the threat a sanction, and the execution of the threat a case of penal force. When the commander is a private person with no authority over the addressee, the command is a naked command, the threat an assault, and the execution of the threat a battery, forcible confinement, or murder, all of which labels indicate that the actions they name are wrongful from the standpoint of a political sovereign claiming a monopoly on rightful coercion. Finally, penal force is either judicial punishment or the imposition of a penalty. It is judicial punishment if it is a reaction by the political sovereign to an antecedent culpable wrong; it is a penalty if, without being attached to a culpable wrong, it is directed toward the achievement of a public goal. By a culpable wrong I mean a wrong deserving of judicial punishment.

Now, the labelling of coercive actions by private actors as "assault," "battery," "murder," etc. from the standpoint of a sovereign's claim of monopoly on rightful coercion would be arbitrary if the sovereign's monopoly simply served its own private ends. For then there would be nothing to distinguish the coercion exercised by the sovereign from that of other private actors beyond the brute fact that the sovereign's claim of monopoly was widely accepted and its self-serving commands generally obeyed, while those who disputed the claim found their commands mostly ignored and their attempts to enforce them resisted. And that fact alone, even if it could distinguish sovereign from nonsovereign, could not distinguish right from wrong. So, it seems reasonable to think that, if coercive action by state officials is to be distinguished from the *wrongful* coercion of private actors, then it must be distinguished most basically by its serving a public interest. Here too definitions are in order. By a public interest I mean one that is shared by all subjects of a political authority bar none; by a private interest I mean one that is possessed by one or some of these subjects and not by others.

That the end of coercive action be a public interest is not, to be sure, sufficient to distinguish official from wrongful coercion. The person exercising force must be legally authorized to act for the public interest, and the force used must be no more than what is reasonably needed to protect the public interest. Yet the concept of a public interest is surely necessary to that distinction, for it is the only kind of interest to whose enforcement against recalcitrants everyone (including the recalcitrant) could rationally assent. Someone who coerces another to serve private ends coerces arbitrarily—that is, in a way that cannot be justified by reasons acceptable to the person coerced; and this holds true whether he coerces only this person now or a multitude into the future under a general command. In the hold-up situation writ large as a political tyranny, the gunman is still a gunman and his victim still a victim.

A public interest might be contingently shared or it might be necessarily shared. A contingently shared common interest is one that individuals possess coincidentally; a necessarily shared interest is one that individuals possess by virtue of their being the kind of beings they are. The problem, of course, with basing coercive authority on a contingently shared interest is that such an authority dissolves in the moment of its assertion, and so is no authority at all. Any willful law-breaker denudes a coercive authority based on a contingently shared interest simply by his defection from that interest. His very challenge to authority transforms the public interest into a private one and therewith penal force into arbitrary force. So let us say that the public interest under which coercive authority is justified must be an interest that is necessarily rather than contingently shared—that is, shared by all subjects just by virtue of their possessing some common nature, whether of living things, animals, rational animals, or of rational agents. I'll call this kind of public interest "public reason", since it is an interest imputed by reason rather than inferred from behaviour or gleaned from self-reports and because it is the fundamental reason or purpose of collective action by people organized in a political society.[1]

The concept of public reason is very abstract, for it is as yet devoid of specific content. What public reason is has to be specified by particular conceptions of the fundamental purpose of political association, of which there are very many possible ones. Because our focus is on penal justice for a liberal society, we can short-circuit what would otherwise be a long discussion by zeroing in immediately on public reason as conceived by the denominations of liberalism. In that way, we bypass instances of public reason given independently of human agency, such as the perpetuation of biological life—its cycles, rhythms, and interdependencies—or

[1] As I use the term, "public reason" is both like and unlike Rawls's notion; see *Political Liberalism* (New York: Columbia University Press, 1993), 212–54. It is like Rawls's in that it refers to a shared normative framework by which individuals with diverse values can unite under a coercive political order. It is unlike Rawls's in that it is not identified with a particular conception of public reason set over against others—namely, the priority of the right. As the bare idea of a fundamental purpose of political association, public reason is hospitable to any plausible conception of a noncontingently shared human interest, including any plausible conception of a fundamental human good.

the ideal development of human nature, or the fullest possible development of human nature as preparation for a supernatural excellence. In contrast to these, liberal conceptions of public reason are conceptions of freedom—of what it means to be a free agent. The denominations of liberalism differ sharply in their understanding of what freedom is and of what the state must do to secure it, but they agree that freedom is the fundamental end of political society.

In the Introduction I distinguished three conceptions of freedom: formal agency, real autonomy, and freedom as citizenship. Ordered to each of these conceptions is a theoretical paradigm for justifying coercive action by officials claiming authority under the conception against those over whom they claim authority. Each paradigm generates a distinctive set of constraints on penal force by which the latter is reconciled with a particular view of the recipient's freedom and so distinguished from wrongful coercion. Since, however, the conception of freedom to which a paradigm is ordered is a partial conception, no paradigm taken alone will yield liability conditions sufficient to distinguish official from arbitrary force. Each will leave a frontier—a kind of administrative badland—where penal policy may roam free of normative restraints. Together, however, the paradigms will generate all the law needed for freedom. The constraints on penal force yielded by the inclusive conception of freedom will comprise necessary and jointly sufficient conditions for distinguishing penal force from wrongful violence. Hence these constraints are conditions of penal liability making up a model general part of a penal law for liberal states.

2. Public Reason as Formal Agency

If we think about why freedom is a plausible conception of public reason—a qualified candidate for the title of fundamental end of public authority—we will arrive at the particular conception of freedom that must form the starting point of any rational derivation of the general conditions of penal liability. Freedom, I suggest, is a plausible conception of public reason because of its emptiness of all interests that may be contingently held. Recall that the move to public reason from a contingently shared common interest was required by the idea of a stable coercive authority. To reach stable ground, therefore, we must abstract from every concrete interest or purpose that may or may not be held by individuals to the one end that is necessarily presupposed in the choice of any end whatever. As Kant taught, that end is the self, understood as a *possibility* or capacity for freely choosing ends and therefore a *potential* for achieving self-authored ends—ends reflecting a deliberative conception of the good. Here, however, it is the bare capacity for free choice that we fix upon, for any exercise of the capacity will once again engage the contingent and idiosyncratic ends from which we sought to abstract.

So isolated, however, the capacity for free choice can be grasped only negatively—as the capacity *not* to be causally determined to seek satisfactions

according to behavioural laws. Because the self *can* reject a caused impulse (hunger, for example) as a motive for action, its following the impulse is something the self freely chooses even when it acts without deliberation or impulsively. Of course, this is not to deny that selves are alive in bodies that are embroiled in cause–effect relationships, nor even to deny that these connections operate with all the appearance of laws when deliberation is absent. What is denied is only that these connections are necessary ones. The possibility of not following an impulse makes following it the choice of a self to whom following the impulse is thus imputable and who is therefore "responsible" for following it. Moreover, as a chooser of its ends, the self is also the final end of action pursuant to its ends—that for the sake of which any action is performed. Because of its negativity and consequent emptiness of content, I'll call this final end formal agency, and I'll call the penal law paradigm it structures the paradigm of formal agency (or the formalist paradigm).

The paradigm of formal agency has as its fundamental end the protection of each agent's capacity to choose the ends of its action against pre-emption, interference, or destruction by the action of another. By pre-emption I mean prevention of the capacity's exercise by applying physical force to an agent so as to restrain or direct its movements according to physical laws. By interference I mean imposing by threat a binary choice upon the agent ("do this or suffer that"), one of which involves a normatively foreclosed acquiescence in wrongful pre-emption or destruction, so that the agent's capacity for choice is normatively pre-empted: it has no legally recognized choice but to obey. By destruction I mean the physical annihilation of the biological conditions (consciousness, life) of the capacity for free choice. All these actions are considered wrongful, not because they challenge the ruler's monopoly of coercion, but because they contradict the end-status of free agency. Were the ruler to pre-empt, interfere with, or destroy an agent's liberty outside its duty to protect the common liberty, it too would have committed a wrong.

Why not happiness?

It might be objected that formal agency is not the only end qualified by its abstractness for the title of public reason. Happiness too is such an end because, although each agent conceives happiness differently according to its subjective likes and dislikes, and although any agreement on what brings happiness would be fortuitous and short-lived, nevertheless happiness as such is necessarily desired by beings who are subjects of desire and are conscious of this. Such beings have not only first-order desires for this or that object but also a second-order desire that their first-order desires be satisfied insofar as satisfying them will yield an optimal level of satisfaction over the long run. The second-order desire is a desire for happiness simply, and it is necessarily shared by all beings who are conscious of themselves as desiring beings. So why begin from formal agency? Why could not punishment be distinguished from wrongful battery or forcible confinement by its being ordered to the greatest possible long-run satisfaction of desire across

all those subject to a political authority? Why should we not say, with Bentham, that punishment is justified if the pain inflicted is necessary to prevent greater pain, and why should we not derive, as he did, liability and exculpatory conditions from this principle?

Despite its abstractness, happiness is not an eligible candidate for public reason. Its disqualifying feature is that, because it consists in the satisfaction of whatever prudent desires people may have, it cannot be pursued independently of seeking to satisfy those desires. As a result, it is prey to the same problem of fragmentation that afflicted a common interest contingently shared and that necessitated the move to public reason in the first place. As soon as the preferences of individuals are such that the satisfaction of some requires the frustration of others, the general happiness can be conceived only as the greatest net happiness, in which sum the sufferings experienced by some people will be justified by the greater gains obtainable by others. Since penal force against someone implies just such a fragmentation of interest, the public interest dissolves into a private one in the very act it was supposed to legitimate as being distinct from wrongful violence.

Formal agency is immune to this kind of disintegration. Because it is an *a priori* capacity to choose ends, formal agency may be respected and protected by a public authority independently of any secondary concern for facilitating the satisfaction of particular ends. Thus exercises of the capacity (for example, in offensive speech) inimical to the general happiness may be protected against interference provided they do not interfere with the capacity for free choice of others, while acts pre-emptive of formal agency (for example, confining people suspected of dangerous tendencies) may be proscribed even if they would yield a net gain of happiness. Unlike happiness, which, though an abstract idea, is still a compendium of satisfactions of subjective ends, formal agency transcends such ends, and this is what qualifies it to be a conception of public reason under which penal force may be distinguished from unjust violence.

The eudaemonist's objection

To this it might be further objected that the very feature of formal agency that recommends it as a conception of public reason disqualifies it as a worthwhile end of political association. Precisely because it transcends material ends, formal agency (it might be argued) is too vacuous an end upon which to base coercive authority. Why would any association take as its fundamental end—as the only end whose claim to satisfaction is indefeasible—something that, as a bare capacity to choose valued ends, is surely only one element of a life worth living? People might desire that their free agency be respected as an essential ingredient of their long-run happiness, but what would be the point of respecting agency in a particular case at the cost of everyone's long-run happiness? Better the happiness of a majority in which everyone may hope to participate than a universal formal agency which no one concerned for the general welfare would absolutize in isolation. Call this the eudaemonist's objection.

The eudaemonist's objection requires two responses. First, though certainly vacuous, formal agency has a special status among the ends forming a person's conception of a worthwhile life. That status derives from its being a condition of choice rather than something itself chosen. We can say that, as the end that is necessarily presupposed in the choice of all ends whose value depends on their being chosen, agency itself is an end whose value is unqualified. This means that, whereas the value of optional ends is relative to the agent whose ends they are, agency's value is absolute or valid for all agents. Moreover, whereas the agent may renounce or compromise the satisfaction of optional ends for the sake of its greatest possible happiness, agency itself is above such trade-offs. This is so because an end whose worth is absolute cannot suitably be treated as one among others, nor can an end that is prior to optional ends be disposed of by the agent as if it were itself optional. Agency, in short, is a locus of unconditioned and incomparable worth, which common parlance renders as "dignity".

So there is point to a public authority's treating agency as an end more fundamental than happiness even though it is formal and empty. The point is to treat agency with the respect due it as a locus of dignity. The way to treat it so is to treat it, in Kant's celebrated formulation "always as an end and never as a means only".[2] And this is why penal force against the agent cannot be justified by the happiness it secures to others, nor can liability and exculpatory conditions be derived from what is necessary for the maximum net happiness. Indeed, it is only because agents bear unqualified worth that their happiness is properly a concern, not only of each severally and of those bound to them by ties of love, but also of a public authority. While the general happiness cannot be the fundamental aim of a sustainable public authority (given its tendency to break up into the happiness of a part), a public authority may facilitate neutrally the pursuit of happiness by free agents whose dignity is what entitles them to such concern. It may do so, that is, subject to an indefeasible constraint of respect for agency. The public authority cannot self-consistently care for the happiness of its subjects at the expense of the very principle that makes the satisfaction of their subjective ends an appropriate object of public concern.

The residual force of the eudaemonist's objection

The foregoing argument, however, is not a complete response to the eudaemonist's objection. This is so because, however compelling the argument for treating agency as a locus of dignity or unconditioned worth, that argument does not establish agency as the *sole* locus of dignity nor, therefore, as the *fundamental* end of political association. The argument for its being fundamental presupposed the relativity of all goods to those who value them and hence the equation of welfare with "happiness",

[2] *Foundations of the Metaphysics of Morals,* trans. LW Beck (Indianapolis: Bobbs-Merrill, 1959), 47.

understood as the prudent satisfaction of subjective ends. If all material ends were subjective values, then formal agency would indeed be the sole unconditioned end and so would be exclusively entitled to the throne of public reason. But observe how all material ends came to be seen as subjective ones. In the negative movement to freedom of choice (as freedom *from* determination), the equation of material ends with biological ends was assumed, and then these ends had to be conceived as optional and subjective for agents rather than as necessary because of their capacity for free choice. Thus the subjectivity of all material ends followed from an initially assumed equation of material ends with biologically given ends; and so formal agency's claim to be the sole unconditioned end is self-contradictorily conditioned by that assumption. Moreover, the assumption seems far-fetched, for surely there are ends—life and good health, for example—that, though given immediately by biology, are valuable mediately through freedom and so valuable to *all* agents because necessary to acting from self-authored ends. If that is so, then the argument for the absolute priority of formal agency does not go through.

I take the latter objection to be unanswerable insofar as it contends against Kant's position that the dignity of formal agency is the fundamental end of coercive authority.[3] Yet the objection overshoots insofar as it views the argument against treating formal agency as *fundamental* as an argument for dismissing it also as an *autonomous* principle for justifying coercive action by public officials. From the fact that formal agency is not the fundamental end of coercive authority it does not follow that it can be immersed in some other fundamental end, such as self-determination or human flourishing; for it may be the case that the paradigm of penal justice ordered separately to formal agency is part of a complex system comprising several instantiations of a comprehensive public reason. Stated otherwise, it may be that the system of penal justice ordered to fundamental public reason is just the unity of the several systems autonomously ordered to constituent conceptions. Such a "federation" of justice paradigms is, after all, a conceptual possibility.

Furthermore, our doubts about whether formal agency is fundamental left its dignity unscathed. For formal agency may be a locus of absolute worth without being the sole such locus provided that plural loci are complementary rather than competitive; and that too is theoretically possible (after all, two agents are loci of mutually compatible claims to absolute worth). But if in rejecting formal agency's claim to be the fundamental end of coercive authority we also reject its claim to be an autonomous end, then we will have indeed denied formal agency's dignity along with its claim to being fundamental even though the latter denial does not entail the former. For, if the capacity to choose ends is valued simply as an ingredient of a common good (autonomy, for example), then the common good will

[3] That this is Kant's position is shown by *The Metaphysics of Morals,* trans. Mary Gregor (Cambridge: Cambridge University Press, 1991) 63: "Freedom (independence from being constrained by another's choice), insofar as it can coexist with the freedom of every other in accordance with a universal law, is the only original right belonging to every man by virtue of his humanity."

determine what respect is due agency, whose worth will thus be conditional on its furthering the good. There will thus be no independent worth to formal agency that an authority charged to promote the good must respect.

The foregoing considerations entitle us to proceed as follows. We begin with the paradigm of penal justice independently ordered to formal agency. We describe the structural features of this paradigm, elucidate its conceptions of right and wrong, and pinpoint the specific difference of criminal wrong. We then present the account of judicial punishment belonging uniquely to this paradigm and derive the liability and exculpatory conditions generated by that account. We also defend this account as one that *partially* reconciles liability to judicial punishment with the freedom and inviolable worth of the agent. Finally, we consider how treating formal agency as the fundamental end of coercive action leads to equivocations regarding the subjectivity of all material ends and permits a liability to noncriminal penalties that contradicts the dignity of agents. This will lead us to a new conception of public reason whose doctrinal offspring will fill out the system of constraints on penal force required by the agent's inviolable worth.

3. The Features of the Formalist Paradigm

The formalism of the formal agency paradigm consists in its isolating as the sole bearer of legal rights the capacity for freely choosing ends. While this capacity is also a potential for achieving self-authored ends, respect is here due the capacity whether or not the potential for achieving self-authored ends is realized or frustrated in the capacity's exercise. This means that there is a coercive duty to respect the capacity even if disrespecting it would better serve the agent's particular ends (happiness) or even long-run autonomy (welfare); that there is a coercive duty *only* to respect the capacity, not a duty to assist the realization of the further potential it carries; and that breach of the duty is a remediable wrong regardless of the beneficent motives of the one breaching. A few examples will bring into focus the difference between exercising the capacity to choose ends and realizing the potential for achieving self-authored ends.

If I act suddenly on an impulse to pocket a gold ring lying under glass in a jewellery store, I have exercised a capacity for free choice inasmuch as I could have rejected the impulse as a motive for action; no law of nature necessitated my behaving as I did. Nevertheless, there is a sense in which I have acted unfreely, because the end of my action was taken up as immediately given by an impulse rather than chosen upon reflection and deliberation. So, while the capacity for free choice was exercised, the potential it carries for acting from self-authored ends was badly realized—indeed hardly realized at all—in the exercise. Similarly, if I am faced with a choice between death by starvation or stealing, any choice I make will exercise my capacity for freely choosing ends, for it can always be said

that I could have chosen otherwise than I did. But my choice will fail to realize my potential for acting from self-authored ends because the only available options are those that have been imposed on me by circumstances and that I would not choose were I empirically able to set ends for myself. Or again, I have a capacity to choose ends, but if my action begets unintended and adverse consequences for me that I could not have foreseen, then my product did not realize the specific ends I chose but was partly the result of extraneous causes owing to which my ends came to nought. Once again, my potential for achieving self-authored ends has been frustrated rather than realized in the exercise of my capacity for choice.

Now, we who are observing the formalist paradigm can see that the distinction between a bare capacity freely to choose ends ("I could have rejected the motive I acted on") and the greater potential it carries ("I can act from a self-authored end") is really one between two conceptions of freedom—a thin one and a more robust one—that we can call, respectively, liberty and real autonomy. But the formalist paradigm does not regard it so. Because it equates material ends with subjective ones having no standing in public reason (and which, apart from a reciprocal bargain, no absolute end could be forced to serve or accommodate), this paradigm must identify the public end with the capacity to choose ends and so make respect for this capacity its sole coercive norm. No distinction is drawn between subjective ends and goods all agents must value as conditions of a realized potential for autonomy; rather, all material ends are considered subjective, so that liberty appears as the only end that free agents could accept as coercively obliging them.

This is not to say, however, that formalism has no place for the idea of autonomy. On the contrary, autonomy is recognized (by Kant, for example) as the richer conception of freedom—the fulfilment of that for which free will is the potential. But autonomy cannot here have the status of a coercive public norm. Because material ends are treated as given biological goals that are subjectively ratified (or not), autonomy must be conceived as a condition one attains by setting aside material ends as motives and by acting purely for the sake of an intellectual end—the right.[4] This, however, makes autonomy a private and inward affair of disinterested motives rather than a coercive public end; after all, no one can be coerced to obey the law without regard to his material interest. So, the only coercive public end is liberty. This fundamental idea—call it the public priority of liberty—is reflected in the following defining features of the formalist paradigm: the equal dignity of agents, the independence of wronging from harming, the exclusion of civil and criminal liability for failing to benefit, and the independence of culpability from evil.

The equal dignity of agents

Because the formalist paradigm views all those with a capacity for choosing ends as bearers of unconditioned worth, it views all agents as equal right-bearers despite

[4] Kant, *Foundations of the Metaphysics of Morals,* 65.

yawning differences in ethical development and regardless of whether the potential for autonomy involved in the capacity is prevented from realization by infirmity or immaturity. Thus persons have rights, not just virtuous or autonomous persons. Career criminals, alcoholics, drug addicts, children, and the mentally ill have the same protection against other agents as everyone else; and whatever exculpatory conditions barring punishment the formalist paradigm prescribes (for example, involuntariness, mistake of fact) apply equally to them—even if their disability or immaturity is the cause of their meeting these conditions and will likely cause nonculpable transgressions again. The schizophrenic who kills a human being whom he perceives as an attacking animal may be nonpenally confined for prophylactic reasons because his disease disables him from realizing his metaphysical potential for achieving self-authored ends (see Chapter 8). But first he will be judged in court as a dignified agent capable of choice and declared not guilty because he chose no wrong.

Not only do all beings for whom free choice is possible have the same right to exercise their capacities for choice; they also have the same legal liabilities. Thus, conditions disabling them from realizing the potential in free choice for autonomous life-shaping do not by themselves disentitle human beings to the respect shown them when they are held legally responsible for their conduct. The mark of responsibility is not an empirical capability for acting autonomously, nor is it (as Duff and Gardner contend) an empirical capability for acting intelligibly from reasons and for being responsive to rational persuasion.[5] These capabilities are more advanced than those required for basic responsibility for conduct. The precondition for responsibility is nothing more than self-conscious agency—awareness of one's freedom of choice, which can exist even when the options one faces are delusional and the delusion fixed beyond argument. Thus infancy negates criminal capacity and so precludes a trial on the facts, but insanity does so only in the rare cases where it vitiates self-consciousness. Otherwise, the insane are tried like everyone else to determine whether, within the world as they saw it, the choices they made were culpable ones. Further, beings with a capacity for free choice cannot be subject to literally irresistible impulses while conscious, though addiction and mental disease might give inclinations a volcanic force to which even a person of reasonable self-control might yield. The formalist paradigm cannot recognize this condition of extreme heteronomy as negating responsibility, though it might admit it as partially excusing for reasons discussed in Chapter 7. In this way the mentally ill and addicted are accorded the dignity of agents who are responsible for their choices however powerful the inclinations they chose to follow.

The foregoing observations permit a glimpse of what would be lost if formalism's flawed equation of liberty with fundamental public reason were taken as a reason for ousting its autonomous legal ordering rather than simply its exclusive

[5] RA Duff, *Answering for Crime* (Oxford: Hart Publishing, 2007), 38–43; John Gardner, *Offences and Defences* (Oxford: Oxford University Press, 2007), 181–5.

ordering. The autonomy of the formalist paradigm alone guarantees the equal end-status of agents and hence the universality of legal rights. This is so because formalism alone sees unqualified worth, not in a moral excellence, but in the morally indiscriminate fact of self-consciousness. Accordingly, were this paradigm merged into one ordered to a richer conception of freedom, rights would be allotted by that conception, and so equality would be lost to an aristocracy of the autonomous. Those whose capacity for free choice cannot be exercised in autonomous or reasoned choices would command not even a defeasible respect for their choices; hence, they could not be wronged either by private actors or by public officials. Like infants, moreover, they would be ineligible for honourable punishment just because of their status—not even fit for judgment as guilty or not guilty—because a material capability for making autonomous choices rather than a capacity for free choice would be the basic precondition of responsibility for conduct.[6] Not only would the chronically heteronomous be vulnerable to coercion for therapeutic and prophylactic reasons (as I argue is justified in Chapter 8); such coercion would be unconstrained by a requirement that their freedom of choice be pre-empted only to the extent needed for public safety; for the capacity for free choice would not be a sufficient basis for rights. Thus they could be held at the sovereign's pleasure even without a determination that they pose a continuing danger. Indeed, since they could not be wronged, even more extreme conclusions could be drawn. Though far from fanciful (parts of this construct can be found in various imperfectly liberal jurisdictions), this picture obviously does not belong in a gallery of liberal legal practices.

The independence of wronging from harming

The formal agency model treats as a wrong any action that manifests disrespect for the capacity to choose ends even if the action inflicts no setback to the victim's ends and even if it actually furthers them. Wrong is simply the pre-emption of, or interference with, the capacity to choose ends.[7] Accordingly, someone who injects medication into an unwilling patient commits a battery even if the medication was essential to his living a productive life, even if the patient desired a productive life, and even if, contrary to what the patient believed, the medication carried no risk of serious side-effects. Conversely, someone who inflicts great harm will be innocent of wrongdoing if he did not interfere with the victim's capacity to choose ends, as in the cases of the business competitor and the tactless speaker, or if the injured person consented to the risk, as in the case of the prize-fighter who kills an opponent in the ring.

[6] See the judgment of McLachlin J in *R v Chaulk* [1990] 3 SCR 1303.

[7] The tort of negligence, I have argued elsewhere, lies at the border of the formalist paradigm, neither wholly in nor wholly out; see Alan Brudner, *The Unity of the Common Law: Studies in Hegelian Jurisprudence* (Berkeley: University of California Press, 1995), 200–1.

That the foregoing examples are commonplaces of the common law of tort and crime should tell us that the "harm principle" of liability, while a favourite of some philosophers, actually plays a very circumscribed role in the penal law. Harm (but not to subjective interests—hence the exclusion of civil liability for pure economic loss) is a condition of civil liability only in the tort of negligence; while in criminal law it plays an ancillary role as an aggravating factor relevant to quantum of punishment and as something whose avoidance might justify what remains a compensable transgression or excuse what remains a culpable transgression. Why harm should be irrelevant in the first instance to criminal liability and yet be secondarily relevant to measure of punishment and defence is a mystery that all single-end theories of the criminal law must find impenetrable. Yet we shall see that it is quite intelligible to the complex whole theory we are expounding.

That harm is irrelevant to wrongdoing in the first instance is explained by the reason for formalism's isolation of the capacity to choose ends as the sole coercive public end. That reason is its equation of goods with subjective values. Because this model identifies material ends with subjectively ratified biological goals, it equates welfare with the satisfaction of subjective ends, which (absent contract) no public authority may coerce absolute ends to serve or accommodate. There is thus no public conception of material interests harm to which could be legally cognizable as an objective harm. Specifically, the formalist model has no conception of welfare as the attainment of goods every reasonable agent must desire as essential to its living an autonomous life. Hence it has no public conception of harm as a setback to these goods. In a context where all material ends are subjective, the only public interest is the capacity to choose ends. And so the only coercive public norm for the formalist paradigm is an injunction to respect this capacity by refraining from pre-empting or interfering with its exercise.

The absence of positive duties

A related feature of the formalist paradigm is that it recognizes as coercible only negative duties not to interfere with the capacity to choose ends; it will not coerce the performance of any positive duties to which the agent has not freely consented in return for binding someone else. By a positive duty I mean a duty to confer a benefit on someone, to act for the sake of his or her welfare. The distinction between negative and positive duties is not, observe, a difference between liability for acts and liability for failing to act. If I fail to remove a car from someone's foot after driving onto it, I have breached a duty not to pre-empt free choice even though what occurred can be indifferently described as an act or an omission.[8] Conversely, if I walk away from someone dying in the street without offering assistance, I have failed to confer a benefit even though I have not been motionless. The difference between negative and positive duties is one between a duty

[8] *Fagan v Commissioner of Metropolitan Police* [1969] 1 QB 439 (CA).

to forbear from pre-empting another agent's capacity for choice and a duty to enhance another's welfare.

The formal agency model recognizes no positive duties, because it makes no distinction between interests relative to individual agents and interests that all human agents have in securing the conditions for living autonomous lives. It equates all interests with subjective interests and all welfare with happiness. And since no one has a right outside of contract to compel an absolute end to serve his or her subjective interests, the agency model will enforce no positive duty of concern.

Still, the absence of coercive positive duties as between private actors is not unique to the formalist paradigm. This feature of formalism will survive into the legal paradigm ordered to real autonomy. This is so because the positive entitlement to the conditions of an autonomous life will be correlative to a prior duty on the public authority to secure these conditions on pain of transformation into an instrument of the economically independent class. Thus a coercive duty on a private actor to promote a stranger's autonomy would amount to his or her unfairly shouldering the public burden, while a coercive duty unilaterally to promote another's happiness would still be inconsistent with his end-status. What is unique to formalism, however, is the absence of coercive positive duties of concern for autonomy even on the citizen. Where the sole public interest is the capacity to exercise choice, the only coercive duty on the citizen is to forbear from interfering with that capacity.

The independence of culpability from evil

Within the formalist paradigm, culpable wrongdoing consists in the choice to interfere with the capacity for choice. Just as the only legally cognizable wrong is interference with the capacity's exercise, so the capacity's exercise is the only legally cognizable criterion for liability to punishment. The wickedness or virtue of the character revealed by the choice is beside the point. This is so because, where all material ends are optional, there is no such thing as natural virtue in the Aristotelian sense of living as the best human would; for natural virtue presupposes that there are nonoptional ends (friendship, citizenship, wisdom) in human nature. But if there is no such thing as natural virtue, then distinctions between virtuous and vicious character themselves reflect subjective judgments having no standing in public reason and so incapable of coercively binding absolute ends. The only publicly acceptable basis of culpability for absolute ends whose ways of life are otherwise beyond public criticism is the choice to interfere with choice. No doubt formalism generates its own alternative to Aristotelian ethics in the ethics of Kantian autonomy. But we have seen why this must be an ethics of motive rather than one of character; and an ethics of disinterested motive cannot coherently be promoted by public coercion (whereas character can be formed by habits ingrained by obedience to good laws). Besides, whether an unlawful act is performed from self-interest or from a maxim of benevolence is a matter hidden from finite agents.

The upshot is that wickedness is neither a necessary nor a sufficient condition of culpability. Thus, it is irrelevant to culpability for wrong that the choice to interfere with choice was out of character for the defendant or that it was prompted by beneficent motives. As the court found, Robert Latimer's killing of his 12-year-old quadriplegic daughter was culpable even if he acted from a loving desire to end her incurable suffering from cerebral palsy.[9] Nor is it relevant to culpability that the choice to interfere with choice was morally blameless because in the circumstances (of duress, say) forbearance would have called for saintliness. As the positive law holds, this is a reason for excusing the culpable, not for exculpating.[10]

Neither is wickedness sufficient for culpability. Thus the wicked intention to harm another is not a culpable intention unless it manifests itself in an external choice to interfere with choice (though a completed interference is unnecessary to an external choice—see Chapter 3), even if there is abundant evidence of the intention and chance alone, not any change of heart, prevented its manifestation. Thus, the would-be arsonist who is interrupted by police while purchasing matches has not committed a culpable attempt even if he confesses his purpose.[11] Moreover, the wicked indifference to another's welfare shown in an unthinking imposition of an obvious risk of serious harm is not the choice to interfere with choice that formalism regards (for reasons that will soon become clear) as alone culpable.

If culpability is independent of wickedness, it follows that punishment cannot be justified as giving the wicked their just deserts. This may come as a surprise to those who associate formalism with Kant and Kant with moral retributivism. Actually, however, Kant is equivocal about what culpability consists in. Sometimes he equates it with wickedness. For example, in one passage in the *Doctrine of Right,* he argues that only the death penalty for murder and rebellion punishes all "in proportion to [their] *inner wickedness*" because the noble would prefer death to enslavement and the base would prefer the opposite.[12] Elsewhere, however, he implies that culpability consists simply in an action whose universalized principle rebounds against the actor. Thus he writes: "whatever undeserved evil you inflict upon another within the people, that you inflict upon yourself. If you insult him, you insult yourself; if you steal from him, you steal from yourself; if you strike him, you strike yourself; if you kill him, you kill yourself."[13]

Hegel is much clearer. In the *Philosophy of Right,* he explicitly distinguishes between the "moral attitude", which takes as essential "the subjective aspect of crime", and the juridical attitude, for which "crime is to be annulled, not because it is the producing of an evil, but because it is the infringement of the right as

[9] *R v Latimer* [2001] 1 SCR 3. [10] See citations in Introduction, n. 10.

[11] Purchasing matches would likely be regarded as merely "preparatory"—too remote from the completed purpose to constitute an attempt.

[12] Kant, *Metaphysics of Morals,* 142–3. Emphasis in original. [13] Ibid, 141.

right . . . ".[14] Elsewhere he says that "[i]n formal right, there is no question of . . . the particular motive behind my volition"[15] and that in the theory of punishment "it is not merely the question of an evil or of this, that, or the other good; the precise point at issue is wrong and the righting of it".[16]

This gives rise to a puzzle. If punishment cannot be justified to absolute ends either as securing the greatest net happiness or as giving the wicked their just deserts, then it cannot be justified in either of the ways proposed by the two most prominent theories of punishment. How then can punishment be justified to absolute ends? Let us begin the story with the formalist account of right and wrong.

4. Formalist Right and Wrong

The formal agency model of penal justice is the system of liability and exculpatory rules organized around a familiar liberal idea: that the fundamental end of coercive authority is to protect the agent's liberty to pursue ends of its choosing insofar as its liberty is compatible with an equal liberty for others. The right to liberty is both a permission to act in all ways that do not interfere with another agent's capacity to act from ends it chooses and an enforceable claim to a remedy against interference by others. Let us first consider the permission and the limits thereof.

The limits reconciling the liberties of separate agents are drawn by property and tort law, whose principles concretize one basic idea: that publicly enforceable permissions to act are individual claims to the sufferance of others that can be recognized without self-abasement by agents equal in dignity to the claimant. That is, legal rights to act certify as suitable for public enforcement individual right-claims that have passed a validity test prior to enforcement; and that test is recognizability by another end. A claim of permission to an exercise of liberty must be confirmable by the agent (or agents) to whom the claim is addressed and who, to give an authentic confirmation, must be capable of remaining independent in recognizing the claim. If accepting the claim would be incompatible with the addressee's end-status, then the claim cannot be objectively validated by an independent other, and so the claim is unsuitable for public enforcement. Of course, there need not be an empirical acceptance by the addressee for a right-claim to be valid, for otherwise the addressee could dictate what the actor can do. It is enough that the addressee *could* accept the claim without compromising his equal worth.

Now, the only claims to another's sufferance that can be recognized by an equal end are those limited by a reciprocal and equal deference to the other's agency. An end can submit to another end without self-loss (and so without disqualifying itself as someone competent to give an effective recognition) if and only if the

[14] GWF Hegel, *Philosophy of Right*, trans. TM Knox (Oxford: Oxford University Press, 1967), para. 99.

[15] Ibid, para. 37. [16] Ibid, para. 99.

other likewise submits to it. Accordingly, that a right-claim to liberty must be recognizable without self-abasement by the agent to whom the claim is addressed does not mean that I cannot hinder or harm another by my actions; after all, to act is potentially to foreclose the actions of others bent on identical objects and to initiate causal chains leading to harm. Rather, the recognizability condition means that hindrances and harms are permissible only under a rule allowing equal and reciprocal hindrances and risks. So, for example, I can disturb my neighbour through the use of my land only as long as I keep within the bounds of ordinary use; and I can accidentally injure another provided that the risk I imposed on him was within the bounds of the socially normal and reciprocal. These legal doctrines instantiate the idea that valid claims to act are those that can be validated by a free and equal other because they are limited by a reciprocal duty to respect his or her liberty. This, then, is the underlying idea of lawful liberty, and I'll call it by various names depending on what seems most natural in the context: the idea of *mutual recognition,* the idea of a *common will,* the intersubjective basis of valid claims of right, the legal norm, or sometimes simply Law.

The idea underlying lawful liberty in the formalist paradigm gives us the essential nature of wrong within that framework. Wrongdoing in general consists in an action whose permissibility cannot be recognized without self-loss by the agent who suffered the effects of the action because the action is incompatible with its equal liberty. If the idea of mutual recognition gives the limit of rightful liberty, then wrongdoing consists in the voluntary transgression of that limit through a choice unrecognizable by an equal. By "voluntary" I do not mean "intentional". A knowing or purposeful transgression is not essential to wrongdoing, because a mistaken belief in permission suffices to challenge the other's rightful boundary and so to require an answering "correction". On the other hand, crossing a boundary through involuntary motion (say, through a push) signifies no challenge to the boundary and thus requires no answering remedy. So, a voluntary crossing of another's rightful boundary without his or her consent is the necessary and sufficient condition of *transgressing* (wrongfully crossing) the boundary.[17]

Wrongdoing in general is particularized as transgressions of the several, more or less intersubjectively constituted, boundaries of the self. That is, while all valid right-claims to action issue from a relation of mutual recognition, not all boundaries of the self limiting others' action are themselves constituted within this relation; some are established nonrelationally or atomistically. The self's body, for example, is a boundary constituted outside of relation to another, so that the boundary is transgressed (the tort of trespass is committed) by any voluntary, nonconsensual contact, even if the defendant reasonably believed there was consent. Here no latitude exists for permissible but unwanted uses (see Chapter 2).[18]

[17] *Weaver v Ward* (1616) 80 ER 284; *Smith v Stone* (1647) Sty 65.

[18] *R v Ewanchuk* [1999] 1 SCR 33. True, action in crowded spaces carries an inevitable risk of jostling, and so there is a presumption of consent from a decision to enter such spaces. But if

Similarly, exclusive possession of things outside the body is constituted outside of relation (by unilateral first possession), and so the boundary is transgressed by a voluntary, unconsented-to encroachment or taking of what is in another's possession; that the trespasser reasonably mistook the thing for his own is irrelevant to his civil liability. By contrast, the landowner's boundary against disturbances of "quiet enjoyment" is itself constituted intersubjectively (because land use inevitably carries externalities), so that the boundary is transgressed, not by any unwanted disturbance, but only by those caused through extraordinary use. Similarly, the self's boundary against harm indirectly caused by another's action is itself a product of mutual recognition, and so the boundary is transgressed, not by any action causing harm, but only by those imposing a socially extraordinary risk. A voluntary, nonconsensual crossing of any of these boundaries is a *right infringement* or wrong, which may be nonculpable (a wrong remediable by compensation but undeserving of judicial punishment) or culpable (a wrong deserving judicial punishment) depending on conditions we shall now discuss.

5. Culpable Wrong and Formalist Punishment

Penal force (arrest, detention, fine, forced service, incarceration) is a right infringement. The official applying force voluntarily and without consent crosses the boundary of another agent's sphere of sovereign choice—the boundary delineated by its body or its property. Penal force is not a prima facie right infringement that an all-things-considered justification erases. It is a right infringement simply. This is so because penal force is not necessary to realize the right of the victim that the wrongdoer violated. If a transgression is threatened, pre-emptive self-defence or the defence of another by official actors realizes the right. If the transgression has already occurred, civil remedies realize the victim's right, for they restore wrongdoer and victim to the normative position of mutual recognition that the wrongdoer's action upset. In that the wrongdoer puts his victim back to where he was, he recants to the victim the claim of superiority his action implied, while the victim gains the wrongdoer's recognition for the right violated. In that sense a civil remedy corrects the injustice. Forcibly repelling the wrongdoer or forcing him to compensate his victim is not a right infringement, for he had no right to perform the action in the first place, and coercing him realizes the right his action challenged; it merely restores the parties to their rightful equilibrium.

However, if action against the wrongdoer's body or property that is *not* in pre-emptive self-defence and *not* compensatory exceeds what is necessary to prevent or correct the wrong to the victim, then that action is itself a right infringement. Since the civil remedy has already (or could) restore wrongdoer and victim to their

someone has clearly indicated nonconsent, there is technically a battery, and it is the doctrine *de minimis non curat lex* that shields the trespasser, not strict right.

rightful equality, any further action against the wrongdoer *sets him back* relative to that position and so is itself a wrong. Thus penal force is a right infringement that must be justified by reasons qualified to do so, not an apparent wrong to be explained away.

The formalist model's equation of the public end with the capacity for choosing ends determines the way it justifies the wrong that judicial punishment is. Since there is within this framework no good necessarily common to agents, penal force cannot be justified in prospective terms as a means to furthering socially desirable ends. Not even security against violence can qualify here as a public interest, since those forming strong alliances will have less to fear from a condition of insecurity than those more isolated; and so they will value additional security less in relation to liberty than those with more to gain from it. Thus, assuming a civil condition already exists guaranteeing rights through the public adjudication of disputes and the enforcement of civil remedies, the strong are likely to see an additional institution of punishment directed to deterrence and the incapacitation of threats as serving the particular interests of the vulnerable. Accordingly, where all good effects are preferences, penal force must be justified independently of its effects, which is to say that it must be justified in terms of formal right alone—as something required by the reality of rights. Given, however, that the victim's liberty right has already (or could) be realized by a civil remedy and that force not justified by preemptive defence or corrective justice is thus itself a right infringement, the requirement that penal force be justified in terms of right alone seems to entail a paradox. It places on the formalist justification of punishment the burden of showing how a right infringement is necessary to the realization of rights.

Criminal desert

The paradox could be resolved if we could identify necessary and sufficient conditions for distinguishing the right infringement that penal force involves from the right infringement of the initial wrongdoer. Where the institution of punishment is justified prospectively by the greatest net happiness, this distinction cannot be made, for on this view officials infringe the recipient's right for the sake of a particular interest (the happiness of the all-but-one they represent), and that is what all wrongdoers do. Moreover, whereas an institution of punishment justified by the general happiness could approve "punishing" someone unconnected to the wrong if doing so were necessary to avoid greater misery to others, one justified by the realization of rights could not. This is so because officials could not realize rights by an action that infringes a right unless there were a logical connection between the antecedent wrong done by the recipient of punishment and the wrong done to him by the punisher; for in that case (and in that case alone) the wrong of punishment would not be something officials do *to* the recipient but would rather be a kind of impersonal nemesis of the wrongdoer's deed that officials execute. Were we able to discern such a connection, therefore, we could say that the right infringement that punishment involves differs from the initial

wrongdoer's in that it is reflexive rather than transitive. It is not something officials do to the recipient but is rather something the recipient brings on himself. Officials perform the act, but because the right infringement is (under conditions I'll discuss shortly) logically implied in the recipient's antecedent wrong, it is not their wrong, and so no remedial action lies against them in turn.

Let us call the connection between the recipient's antecedent wrong and the wrong he suffers "desert". Now, of course, there are many possible views as to what actions or psychological states deserve a punitive response of some kind. One may think, for example, that an adverse consequence is deserved by someone who acts immorally by being verbally abusive to a spouse, who harbours an immoral wish for a competitor, or who repeatedly causes harm to others through selfish inattention and insensitivity. All these evils deserve natural punishment (of which bad conscience is a form), but our sole interest is in judicial punishment, and the formalist paradigm dictates one particular understanding of what it means for punishment of this kind to be deserved.

Suppose Surgeon (S), while operating on Patient (P) to remove a cancerous prostate, also removes a healthy appendix, erroneously believing he has P's consent harmlessly to remove a potential source of trouble. S has wronged P, because he has pre-empted P's capacity to direct his body according to his own choice. He will thus have to pay P at least nominal damages for his trespass. Because, however, S's interference with P's freedom of choice was unwitting, it signified no denial of P's capacity for rights, hence no denial that agency is that capacity. Rather, it merely expressed a mistaken belief that P had, in exercising his capacity for choice, given S permission to remove his appendix. Therefore, S's transgression was limited to a transgression against P: it had no more general significance. It is therefore for P to decide whether he wants to assert his right against S; and once he has been fully compensated for the infringement, nothing remains to be done to vindicate the right, that is, to realize it in the face of an action inconsistent with its reality. Even if P decides not to pursue S, P's right is intact, for S is off the hook only by the grace of P—only by virtue of P's voluntary decision to forgo his right of action. Since, moreover, P's right can be fully realized by S's compensating him (and since P's right alone has been infringed), any coercive action taken against S not justified by a compensatory rationale is a gratuitous piece of violence by which S is unjustifiably wronged.[19]

Let's now vary the facts somewhat. Suppose that S removes P's appendix knowing that P has refused consent to any procedure not specifically authorized in advance. S has obviously still transgressed a boundary demarcating a sphere wherein P's choices rightfully rule, and so S will still have to compensate P. However, S's action now bears a significance transcending his infringement of P's right. Because S *knowingly* pre-empted P's capacity to choose for his body,

[19] I'll consider in a moment the possibility that penal force is justified by the need to deter those who would otherwise treat compensation as a worthwhile price for their wrongful activity. But the wrongdoer in our example has not done this, and so to punish him for that purpose would be to use him as a means only.

substituting his own choice for P's, his action signified a denial of P's end-status and so a denial that P is a right bearer. True, this is a specific insult to P's dignity requiring a compensatory award that reflects this additional dimension of S's private wrong. However, S's wrong is now not limited to a wrong against P. This is so because P, whose right-bearing status S denied, is an agent with a capacity for freely choosing ends. S may not have known this about P, but if P is functioning minimally as a specimen of *homo sapiens,* the thinking Agent in S's shoes will infer a capacity for choosing ends, and this inference can be imputed to S unless he delusionally mistook P for a nonhuman thing (which he did not). Since S can be taken to have assumed P was an agent, his intentional interference with P's body signified a denial that *agents* have rights to choose for their bodies just by virtue of their free will and that all agents are thus equal in dignity.[20] This, however, is tantamount to a denial of the possibility of rights, for (as we have seen) a claim of right to others' sufferance of one's action remains a subjective assertion unless it can obtain a validating recognition from the persons sought to be bound; and validation can come only from independent ends who can bind reciprocally. Rights can exist objectively only among free and equal agents. Accordingly, S's intentional pre-emption of P's capacity for choice signified a denial of the possibility of valid rights and therewith a claimed permission to an unlimited liberty.[21]

In the previous scenario, no such implication followed from S's action. There, on the contrary, S acknowledged the existence of a boundary to his liberty; he simply erred in thinking he had not crossed it. In this case, however, S either denies that boundaries to liberty exist (claims that each agent may do as it pleases) or he claims that, while existing for others, they do not bind him. In either case, he lays claim to an unlimited liberty. Whether such a claim actually crossed S's mind is irrelevant, for what counts is what the thinking Agent standing in S's shoes meant by S's choice to remove P's appendix without P's consent. The thinking Agent will draw the logical implication of an intentional interference with P's body, so that even if S merely excepted himself from the limits applying to others, the Agent will generalize the self-exemption to everyone who may similarly assert his or her special worth and conclude that rights do not exist. And that denial of rights can be imputed to S (the empirical agent) as a necessary implication of his choice to remove P's appendix against P's will.

But more than that can be imputed to S. If the thinking Agent in S's shoes denied the existence of rights to spheres of free choice in general, then it (the

[20] In crimes motivated by disdain for a group, this denial is conscious and explicit. Thus, someone may attack another believing that humanity is defined, not by agency, but by race, that only a few have rights, and that his victim is not among them.

[21] Thus Arthur Ripstein errs in thinking that the essence of crime is the denial of the *victim's* right rather than of rights simply; see *Equality, Responsibility, and the Law* (Cambridge: Cambridge University Press, 1999), 149. This error stems from Ripstein's tort-centred view of crime, according to which crime is a species of the genus tort differentiated by intention; 147–63. Jean Hampton fell into the same error; see 'An Expressive Theory of Retribution,' in W Cragg, ed., *Retributivism and its Critics* (Stuttgart: Steiner, 1992), 1–25. For a view of crime as *sui generis,* see Chapter 3 below.

Agent) by necessary implication also denied them for itself; and that denial too can be imputed to S just by virtue of his being an agent with a capacity to think. Moreover, given his implicit denial of his own right to liberty, S cannot justly complain if his freedom of choice is pre-empted in turn, for he has implicitly assented to this; he affirmed the right-denying principle ("I assert a permission to act unlimited by another's agency") by which his own liberty is pre-empted.[22] True, his denial of rights was mistaken, for agency is the bearer of an unqualified worth which no agent can claim without also recognizing it in others. This is why, despite S's implicit right denial, penal force against him is still an infringement of *his* right and why his implicit assent to being handcuffed, shackled, confined, etc., is assent to actions that *are* transgressions. It would thus be incorrect to say that S has forfeited his right to liberty. He has, however, forfeited his right to complain of a right infringement by virtue of his own (implicit) right denial. Accordingly, if S's right to liberty is indeed violated in turn, he may be said to have deserved this in the specific sense of desert understood by the formal agency framework. That is, he called it upon himself. He implicitly authorized it, so that there is no element of an external or transitive imposition in the penal transgression.

Observe that this conception of desert does not refer to a connection between an indeterminate evil (blameworthy deed, omission, or disposition) and an indeterminate detriment expressing blame, censure, condemnation, or whatever. Rather, it refers to a connection between a certain kind of right infringement (one in which a right denial is implicit) and judicial punishment, specifying when a *right infringement* is deserved. Thus, no matter how deserving of suffering an evil character may be, he does not deserve a right infringement unless he deserves it in the tight sense demanded by the formalist paradigm. That is, we must be able to say that actions amounting to an infringement of the wrongdoer's right are authorized by the right denial either explicit or implicit in his or her antecedent choice. So penal force is deserved if the recipient:

(a) interfered with agency with the aim of demonstrating the nonexistence of rights;
(b) interfered with agency in the belief that this human being had no rights or belonged to a group whose members were right-less;
(c) knowingly interfered with agency or knowingly took the risk of doing so (S's case); or
(d) performed actions short of actually interfering with agency but unambiguously manifesting an intention to interfere (see Chapter 3).

But if penal force is deserved under the condition that it be authorized by the practical right denial of the recipient, then deserved punishment is compatible with freedom and distinguishable from the initial wrongdoer's wrong.

[22] Why assertion without externalization is not enough is discussed in Chapter 3.

The point of punishing: a Kantian suggestion

So far, however, we have shown only how penal force can be reconciled with freedom *if* it is applied in response to an intentional interference. We have not yet shown why it generally ought to be applied. Failing a demonstration of purpose, however, we will have come up short in showing how punishment can be compatible with freedom and distinguished from the wrongdoer's violence. This is so because, if there is no end of punishment admissible by formalist public reason, then there is a logical gap between the crime and the punitive response such that the latter cannot be taken to have been strictly self-imposed. What is so far entailed by the criminal's intentional interference with agency is only his or her moral vulnerability to coercion, not the coercion itself. Crime gives the state a licence to apply force to the criminal but does not, as so far explained, demand use of the licence. But if the state has an absolute discretion to punish or not, then its arbitrary choice to punish cannot be conceived as self-willed by the recipient, and so it retains an element of external violence. Accordingly, we must show how the administration of punishment is itself self-willed by the recipient; and we do this by showing that there is a public reason to punish crimes as a general rule—one that the thinking Agent will endorse for the sake of its freedom and whose endorsement by the thinking Agent can be imputed to the empirical one.

Now it might seem that the formalist paradigm must falter here. Since it treats all material ends as subjective, formalism cannot regard the beneficial effects of punishment as a public reason to punish endorsable by the thinking Agent. Hence it cannot subscribe to a "mixed" theory of punishment according to which the latter is generally justified by socially desirable ends but can be imposed in the particular case only if the recipient deserves to be punished.[23] For formalism, desert cannot function simply as a constraint on punishing for ends unrelated to desert; rather, desert must somehow figure in the reason for punishing.

Does this mean that the only reason formalism can give for punishing is to give wrongdoers their just deserts? Indeed, most critics of the retributive theory of punishment take this as its reason for punishing and then have an easy time discrediting the theory because of the way it ties punishment to passion—to moral hatred, anger, vengefulness, and *hubris*.[24] However, the critics attack not retributivism, but *moral* retributivism. The retributivism generated by the formalist paradigm cannot regard meting out just deserts for its own sake as the point of punishment either; for we have just seen that crime entails only the coercibility of the criminal, not coercion itself, so that to give criminals what they deserve even after they have fully compensated their victims is an undetermined decision lacking a public purpose the agent could endorse. "Because they deserve it" cannot be

[23] HLA Hart, *Punishment and Responsibility* (Oxford: Clarendon Press, 1968), 1–13.
[24] Ibid, 169–73.

an adequate answer to "Why punish the guilty?" because desert does not necessitate punishment.

If formalism can regard neither the beneficial effects of punishment nor the meting out of just deserts for its own sake as its public reason for punishing, what purpose can it recognize? One suggestion is offered by Arthur Ripstein by way of an interpretation of Kant's view of punishment.[25] According to Ripstein's Kant, the point of legal punishment is to secure a "public rightful condition" (or what we would call the rule of Law) by deterring would-be strategic wrongdoers with the threat of having their wrongs visited back upon them. Were civil damages the sole remedy for a right infringement, those inclined to wrongdoing could treat damages as a worthwhile cost of pursuing their aims. They could thus "make the existence of a public rightful order the means through which they wrong others".[26] For Ripstein, this means that a legal order confined to civil remedies would "authorize" wrongdoing provided damages were paid. Payment of damages "could in principle *entitle* the criminal to wrong his victim, as a matter of right, simply by paying the requisite fee".[27] However, no self-respecting agent would acknowledge a duty to renounce self-help according to his private determination of right and wrong to an order that failed to actualize rights publicly. Accordingly, a threat of punishment for strategic wrongdoing is required to make such wrongdoing "normatively unavailable" so as to produce a public rightful condition. This is a public purpose consistent with the subjectivity of ends that the thinking Agent will endorse.

Alas, this argument fails. It cannot be true that wrongdoers' treating damages as a fee for wrongdoing *actually* makes damages a fee. From the fact that they can do so with impunity it does not follow that they cannot wrong their victims, provided there is a public system for adjudicating disputes and for civilly enforcing rights. If, as tort theory tells us, compensation for damages (including for the lost choice to alienate) corrects the defendant's injustice to the plaintiff, and if a system for enforcing court-ordered civil remedies exists, then rights exist in a civil order prior to punishment. Indeed, if rights did not precede the practice of punishment, all self-styled punishment would be an arrogation of punitive authority to enforce a private opinion of right. Accordingly, strategic wrongdoing is already normatively unavailable in a civil order, whether or not it is practically unavailable. The fact that wrongdoers can use the civil order for their purposes without hindrance means that freedom is insecure; it does not mean that rights are nonexistent. But if a civil order confined to enforceable civil remedies is one in which rights have crystallized, then it is one to which self-respecting agents can renounce their right of self-help.

Perhaps Ripstein means that, in a legal order without punishment, the strategic wrongdoer can wrong his victim but cannot do wrong by paying damages

[25] '*In Extremis*,' (2005) 2 *Ohio State Journal of Criminal Law*, 415.
[26] Ibid, at 417. [27] Ibid, my emphasis.

as a price of wrongdoing. Hence *compensated* wrongdoing is permitted. Put this way, however, the argument does not take us to punishment for crime generally. The injunctive remedy backed by a threat of punishment for contempt of court is a forward line of defence making strategic wrongdoing normatively unavailable. If an injunction is unfeasible, strategic wrongdoing is rendered normatively unavailable by publicly invalidating the gains made from it. This is done by requiring the strategic wrongdoer to disgorge his profit (or its monetary equivalent if the profit is psychological); it is not necessary to punish him.[28] True, wiping out profits from strategic transgressions is often the point of what judges call "punitive damages", but they are punitive only in the sense that they exceed what is needed to compensate the victim; they retain the nature of the tort remedy in that they merely restore the wrongdoer to his position before the wrongful transaction (perhaps this is why the criminal standard of proof is thought to be unnecessary).[29] To punish in the strict sense, however, is to set the recipient back from the *status quo ante,* for at that position he was in enjoyment of his rights, whereas punishment violates them. Punitive damages, while they take away a gain from the defendant that does not belong to the plaintiff, do not violate the strategic wrongdoer's rights, for he had no right to a gain made from an invalid exercise of liberty.[30] Accordingly, since the forced disgorgement of ill-gotten gains publicly invalidates strategic transgressions, punishing the transgressor is not yet determined in the way that it must be if it is to be conceivable as self-imposed.

We could try to repair Ripstein's argument by saying that, because a legal order confined to civil remedies (including disgorgement) could not *deter* strategic wrongdoing, it could not adequately protect persons in their freedom; hence, self-respecting persons could not surrender their right of self-help to it. What is needed for a rightful condition is a threat of punishment for strategic wrongdoing sufficient to give everyone confidence in the public protection of their liberty. Though reminiscent of Hobbes, this argument appeals to dignity rather than felicity, for without a rightful condition, no one can wrong another by enforcing his own opinion of right.

Even as thus restated, however, the argument misconceives the basis of the duty to sustain a rightful condition as Kant himself understood it. The fact that, without an institution of punishment to supplement civil enforcement, there would be

[28] See *Edwards v Lee's Administrator* 265 Ky 418 (1936). It is immaterial that the wrongdoer may expect to consume his gains and be otherwise judgment-proof; for this goes to his idiosyncratic amenability to dissuasion, not to the systemic normative availability of strategic wrongdoing. After all, any particular strategic wrongdoer may also happen to enjoy the security of prison life.

[29] *Rookes v Barnard* [1964] AC 1129 (HL).

[30] It is because Ripstein fails to see the nonequivalence between punitive damages of this sort and punishment that he thinks the justification for the former can serve as an adequate justification of the latter; *Equality, Responsibility, and the Law,* 149–55. Of course, punitive damages might be assessed to deter rather than simply to remove ill-gotten gains, in which case they would be punitive in the true sense of the word. But the unanswered question is how the deterrent could be necessary to render strategic wrongdoing *normatively* unavailable so as to make a rightful condition possible. If a rightful condition did not precede punishment, what would punishment be for?

no assured disincentive to strategic wrongdoing cannot absolve from the duty to renounce self-help, for that duty is not conditional on the low probability of criminal events. Rather, it is conditional on the continuing existence of public institutions for adjudicating disputes and enforcing private rights.[31] Thus, if there exists a full range of civil remedies including injunctions, compensation for lost choice to alienate, and disgorgement of gains from strategic wrongdoing, and if a credible system for enforcing court-ordered civil remedies (through evictions, seizure of assets, penalties for contempt of court, etc.) exists, then a rightful condition exists and so too a perfect duty not to defect from it. The threat of punishment may then be necessary to deter intentional wrongdoing for reasons of public safety, psychological security, etc.; but it cannot be necessary to *make sense* of wrongdoing.

Moreover, the thinking Agent cannot accept an institution of punishment conceived in its nature as a threat aimed at securing people against criminal acts. Though empirical agents may or may not need the threat of punishment to deter them from strategic wrongdoing, the thinking Agent most certainly does not: having no subjective ends, it forbears from wrong (wills the right) for the sake of its dignity alone. Indeed, to the thinking Agent, punishment conceived as by nature an external threat is an affront to dignity, for, like any threat to freedom, it treats agents as manipulable objects rather than as ends. Though the thinking Agent will concede that its motive for forbearance from wrong cannot be relied upon by the civil order and that disincentives to wrongdoing are needed for some or even most, it will not allow that punishment be conceived *essentially* in this way; for if it is, then punishment is incompatible with its freedom. Rather, it will admit only that punishment is a threat relatively or contingently—that is, for those who happen to be inclined to wrongdoing whenever they can derive advantage from it. For itself, the thinking Agent will demand a justification for punishing that appeals to its interest in dignity without relying on the fear of disadvantage (for it has none).

The point of punishing: Hegel's solution

So what public purpose acceptable to the thinking Agent can the formal agency paradigm assign to punishment? In paragraph 97 of the *Philosophy of Right,* Hegel offers an answer:

> The infringement of right as right is something that has positive existence in the external world, though inherently it is nothing at all. The manifestation of its nullity is the appearance, also in the external world, of the annihilation of the infringement. This is the right actualized, the necessity of the right mediating itself with itself by annulling what has infringed it.

We can explain this terse passage by returning to the case of the Arrogant Surgeon. S removes P's appendix with knowledge of P's express choice that he not do so.

[31] Kant, *The Metaphysics of Morals,* 120–4.

We saw that this transgression signified S's denial of P's end-status *qua* agent and therewith a denial of the possibility of rights, which can exist objectively only among mutually recognizing ends. S's implicit denial of rights in general was also an implicit claim of permission to an unlimited liberty, for if there are no rights to respect for liberty, then S has an unlimited permission to act as he pleases, as does everyone else. However, a claim of permission to an unlimited liberty is self-contradictory, for its corollary when generalized is that no one, including S, may complain if someone pre-empts his liberty by physical force. Thus S's claim of permission to an unlimited liberty—a claim implicit in his intentional pre-emption of another agent's choice—is self-destroying. The public point of punishment is to manifest this aporetic result of S's claim by pre-empting S's liberty and, in doing so, to vindicate the truth (denied by S) that the only valid claims of permission to act are those that can be validated by equal ends. The thinking Agent will endorse an institution of punishment directed to this purpose, for it is one by which its right to lawful liberty is vindicated against a claim that, had it been allowed to stand, would have signified the right's nonexistence.

Observe that this is a reason for punishing that invokes neither material ends nor the bare fact of the wrongdoer's desert. The point of punishing is to realize everyone's right to act for any material end against an intentional interloper's denial of rights. Thus the criminal wrongdoer's desert—his practical denial of rights—while not a sufficient reason to punish, figures necessarily in the sufficient reason; for the vindication of mutual recognition (Law) as the sole ground of valid right-claims presupposes a challenge to its being this ground and consists precisely in manifesting the self-destructive result of the challenge. That the wrongdoer has authorized his own coercion by his claim of permission to an unlimited liberty shows that the only coherent claims to liberty are those limited by mutual recognition. Thus the right-vindicating purpose of punishment can be coherently furthered only against someone who deserves to be punished.

Neither compensation nor punitive damages suffices to invalidate the right-denying claim expressed by an intentional interference. Compensation adequately realizes P's right to choose his ends against S's mistaken claim that P is not a right bearer; but it fails to address the wider implication of S's intentional interference, namely, that (since P is an agent) there can be valid claims to act outside the bounds of mutual respect for agency. Moreover, to require only that S compensate P and that S disgorge his gain from the wrongful transaction merely restores S to the *status quo ante* where he enjoyed a right to mutually respectful liberty; whereas his claim is just that the *status quo ante* is invalid—that he has a permission to act unlimited by respect for other agents. The only remedial act qualified to answer that claim is one that listens to its challenge to the normativity of mutual respect, that then manifests the self-contradictoriness of the claim in the wrongdoer's own vulnerability to coercion, and that thereby vindicates mutual recognition as the sole ground of valid right claims.

It might be objected that all this can take place in a philosophy seminar. If the aim of an institutional response to crime is to vindicate mutual respect as the ground of rights by showing the self-contradictoriness of claims of right to nonrespectful liberty, why must the response take the form of punishment? Why not teach the truth about rights in school? If the wrongdoer's principle is a logical nullity, after all, then rights are invincible and require no conquest over a pitiful challenger to prove their strength. They can turn the other cheek. Moreover, if the application of punishment is not necessary to achieve this rarefied purpose, is it not a piece of gratuitous violence indistinguishable from the wrongdoer's?

Perhaps the lesson taught by punishment could be taught in school if the intentional interloper asserted his claim in the classroom. But he did not. He interfered with someone's agency, and his interference, because of its intentionality, expressed a right-denying principle. By *actualizing* a right-denying principle, however, the wrongdoer gave that principle an appearance of worldly authority, of existential force. He gave it, that is, the appearance of a law. That appearance would be allowed to stand if the nemesis of his principle were demonstrated by speech alone, for what is true in logic might not be true in the world; in the world, after all, there has been a violation expressing a claim to an absolute liberty—a claim that the civil remedy did not refute. So, by visiting the self-destructive consequence of the wrongdoer's principle upon him, punishment removes the appearance of its worldly validity and vindicates the worldly authority of Law.

One might further object that the authority of Law is inherently invincible and cannot be negated by logically absurd claims making pretences to existential validity. This is true. However, the invincibility of Law is a reason only for denying that the state *must* punish in all cases of intentional wrongdoing; it is not a reason for denying that the state ought to punish as a general rule. For if it did not, Law's authority would remain pervasively unrealized and hence no authority at all. Were private damages the sole remedy for an intentional interference, the authority of Law as mutual respect for liberty would remain deficient in two respects: first, the *existential* authority of mutual respect would depend on the plaintiff's contingent ability to sue; second (even if he did sue), the defendant's challenge to the *normative* authority of mutual respect as the sole ground of valid right claims would go unanswered because the civil suit takes that authority for granted. Corrective justice restores the parties to equality without addressing the intentional interloper's implicit denial that equality is Law.[32]

Of course, the foregoing argument need not deny that punishment may be used as a deterrent threat against strategic wrongdoing provided that its application is otherwise consistent with self-imposition. Nor need it deny that deterrence is the point of punishing from the everyday perspective of the empirical agent,

[32] In that corrective justice restores the parties to a *status quo ante* that it assumes is valid, whereas retributive justice sets the wrongdoer back from the *status quo ante* in order to demonstrate its validity, retributive justice is not simply a form of corrective justice. It must be regarded as *sui generis,* so that, *pace* Aristotle, there are three forms of justice, not two.

who must be assumed to act for the sake of self-interest, who therefore can be persuaded to forgo the advantage of crime only by the threat of greater disadvantage, and for whom it would be irrational to surrender its right of self-help to an authority no one would rationally fear. The point, however, is that deterrence cannot be the liberal *justification* of punishment, since the thinking Agent cannot consistently with its end-status endorse an institution conceived essentially as threat even if the threat were *empirically* required to secure a lawful condition; for that would make freedom dependent on unfreedom. The only reason it is permissible for the state to give criminals what they deserve so that others will be deterred is that punishment is independently justified to the thinking Agent as the practical annulment of a claim of absolute freedom that vindicates the normative authority of lawful freedom.

6. A Response to Critics of Retributivism

The theory of punishment I have presented so far may be called classical legal retributivism. It is classical in that it stems from the writings of the great philosophers who, present at the birth of the liberal *Rechtsstaat,* first articulated (though differently) its theoretical foundations. I mean, of course, Hegel's *Philosophy of Right* and Kant's *Doctrine of Right.* It is *legal* retributivism in that it justifies judicial punishment as vindicating the authority of Law rather than as giving evil its just deserts. How does this version of retributivism fare against the most common objections to retributivism?

The most common criticisms presuppose a certain portrait of the adversary. According to this description, retributivism is a deontological rather than a teleological theory of punishment in that it justifies punishment as a morally obligatory response to an antecedent evil rather than as a means for attaining socially desirable ends. More specifically, retributivism is (in the eyes of its critics) the thesis that punishment is justified by the moral duty to give those whose actions evince moral fault their just deserts and that the performance of this duty has intrinsic value quite apart from the consequential benefits punishment may yield in reducing crime. There are two parts to this thesis: first, people ought to be punished if and only if they deserve it; second, people ought to be punished with the severity and only with the severity they deserve.

The moral duty to punish evildoers

The portrait thus painted is of moral retributivism. Its critics commonly distinguish between a strong and a weak variant but typically regard the strong version as the only self-consistently deontological one.[33] Strong moral retributivism

[33] See Russ Shafer-Landau, 'The Failure of Retributivism,' in J Feinberg and J Coleman, eds., *Philosophy of Law*, 7th ed. (Belmont, Cal: Thomson, Wadsworth, 2004), 837–8.

argues that *all* those deserving of punishment ought to be punished with the severity they deserve because and only because they deserve it; weak moral retributivism holds that *only* the deserving *may* be punished with a severity *no greater* than they deserve. Critics regard the strong version as the only coherent retributivism, because if desert functions as a constraint on, rather than as a reason for, punishing, then forward-looking considerations tell us why to punish and, subject to the ceiling, how much. But if that is so, why should these considerations not also give us the reason for confining punishment to the deserving and so determine when that constraint may be exceptionally relaxed?

Yet strong moral retributivism is open to objections whose undoubted force has always made consequentialist theories of punishment look attractive by comparison. Take the first part of the "just deserts" thesis—that the wicked ought to be punished because and only because they deserve it. Here one might object that, unless the state is to enforce the whole of morality, a distinction must be drawn between criminal and noncriminal immorality. Yet strong moral retributivism by itself yields no criterion for this distinction. A "public harm" principle might draw the line but not without transforming strong retributivism into weak retributivism, for now judicial punishment in particular would be justified by harm prevention. But if no moral purpose can, compatibly with strong retributivism, separate criminal from noncriminal immorality, then moral anger as expressed in politics must do so; and so to make wickedness the object of punishment is to coerce the convict to satisfy the moral passion of others.[34] Retributivism, on this view, becomes the executive arm of legal moralism—the view that it is permissible to use the criminal law to enforce the socially dominant opinion as to which evils are most deserving of public retribution. But this view is illiberal because inconsistent with agent inviolability, which forbids coercing the agent to serve others' ends.

The logical gap between moral retributivism and state punishment causes a problem of a related sort. If we ask, "but why should the community give evildoers what they deserve?" the answer must be that its doing so is intrinsically valuable and hence self-justifying. There is no need for a justification to punish evil; that the agent is an evildoer is sufficient reason. But even if it is intrinsically valuable that the morally wicked receive the punishment they deserve, it does not follow that the state should do the punishing. Since other options are open (social censure, shunning, blacklisting, leaving to divine justice, for example), *judicial* punishment of the wicked requires a further justification referring to some end extrinsic to retribution—for example, the moral unity of the political community or the satisfaction of the community's moral indignation—to be achieved by the state's assuming the role of moral judge and enforcer. Yet once such ends are invoked, the whole of political prudence is also engaged; and so one must inquire whether these conjectural benefits of state punishment would justify the high cost of a police force, criminal courts, and prisons were crime reduction not

[34] James Stephen, *A History of the Criminal Law of England* (Macmillan: London, 1883), vol. II, 81; for a critique, see Hart, *Punishment and Responsibility,* 169–73.

also a consequence of legal punishment.[35] It would seem that, to obtain the state's sword, the moral retributivist must make a Faustian bargain.

Finally, critics point out that many actions involving no or trivial moral fault (such as unwittingly selling alcohol to a minor) are penalized for the sake of public goals, and no one suggests there is anything wrong with this. Thus, if evildoing is not a sufficient condition of *state* punishment, neither is it a necessary one.[36]

By now it should be clear that legal retributivism is untouched by criticisms such as these. First, legal retributivism nowhere argues that the reason for punishing is to give those whose actions evince moral fault their just deserts. From the broad class of evildoers, it singles out those whose actions manifest choices to which a denial of rights must be imputed, and it selects only these for retributive punishment. Thus, for legal retributivism, moral wickedness is not a sufficient condition for eligibility for judicial punishment, and so there is no need for politics or an extrinsic end to make the final selection; legal retributivism itself determines which wrongs are criminal. Nor is moral wickedness a necessary condition of criminal liability. The physician who, from a benevolent impulse (not amounting to the vice of excessive beneficence), injects a nonconsenting friend with life-saving and riskless medication is, for legal retributivism, a right denier deserving of punishment. Accordingly, judicial punishment is the executive arm, not of legal moralism, but of equal agency rights; and the convict is subject, not to moral opinion, but to the public reason of mutual recognition.

Second, legal retributivism nowhere claims that punishing those deserving of punishment is justified for its own sake. Not even Kant, whose famous plea for executing the last murderer in a society doomed the next day to extinction is usually invoked as the representative statement of this thesis, held this view. Rather, Kant vacillated between two views—one Hobbesian, the other Hegelian. The fact that he regarded a murder-or-die dilemma as robbing punishment of its justificatory point (because the threat of an imminent death defangs the threat of a future one) suggests he thought its justifying point was to deter crime so as to support the state's monopoly of coercion constitutive of a civil condition.[37] That is his Hobbesian side. However, his call for the execution of the last murderer in a moribund civil condition is inconsistent with that view and more consonant with his competing view that punishing the deserving is required by justice "as the Idea of judicial authority"—that is, as the realization of the authority of Law.[38] This is Hegel's univocal view. The point of punishing, Hegel writes, "is to annul the crime, which otherwise would have been held valid, and to restore the right".[39]

Such a justification eludes the deontological/teleological/mixed classification of punishment theories as well as the strong/weak classification of retributive theories, showing how simplistic these categories are. Because legal retributivism

[35] Shafer-Landau, 'The Failure of Retributivism,' 834.
[36] Ibid, 832.
[37] Kant, *The Metaphysics of Morals,* 60–61.
[38] Ibid, 143.
[39] *Philosophy of Right,* para. 99.

justifies punishment as vindicating the normative authority of mutual recognition, it can be called teleological; but because its end is right vindication rather than crime prevention or social welfare more generally, it may also be called deontological as opposed to eudaemonist or consequentialist. So, legal retributivism does not sell its deontological soul for the state's goal-directed sword. Yet it is not mixed either, because, viewing punishment as a denial of a denial, legal retributivism sees a necessary connection between the point of punishing and the restriction of punishment to the deserving. The authority of Law requires vindication only against one who denied that authority. Further, because legal retributivism regards desert as a constraint on, rather than a reason for, punishing, it appears to fall in the weak category. But because its reason for punishing is to deny a practical denial of rights that would appear valid were its self-contradictoriness not practically demonstrated, it holds that the state ought generally to punish those who have licensed their coercion whether or not doing so would prevent crime, and so it may also be called strong. We can say, then, that legal retributivism reconciles the view that desert is a constraint on, rather than a reason for, punishing with a nonconsequentialist view of the purpose of punishment. It thus shows how punishment can be justified by an end without losing its essential connection to desert.

Finally, the fact that the state commonly penalizes action inimical to the public welfare but evincing no moral fault does not tell against legal retributivism, for the latter's account of punishment presupposes crimes involving infringements of agency rights and so has application only to that sphere. Legal retributivism makes no claim to being a comprehensive theory of penal force; rather, it is a theory only of punishment for criminal right-infringements. How legal retributivism can be coherently integrated into a broader theory of penal force that covers public welfare offences is the subject of Chapter 5.

The moral duty to fit the punishment to the crime

Consider now the standard criticisms of retributivism understood as a theory prescribing a method for apportioning punishment to crimes. Here the usual objections run as follows. To punish according to desert is to make the punishment "fit" the crime—that is, to make it somehow equal to the crime. The only certain equation between crime and punishment, however, is an identity in kind between the physical injury the wrongdoer has inflicted and the physical injury done to the wrongdoer in return: an eye for an eye, a whipping for an assault, death for a homicide. But doing to wrongdoers what they have done to their victims ignores the possibility that the same injury may have different welfare impacts on different people. Suppose, for example, that the victim had only one eye and the wrongdoer has two, or that the wrongdoer was blind to begin with. This difficulty might lead us to think that, if equality with the crime is the measure of deserved punishment, what must be equalized is the suffering experienced by the victim and the wrongdoer. But interpersonal comparisons of suffering

are impossible for an external judge, and, in any case, how could a schedule of equivalent sufferings be determined in advance by a general law applying to everyone? Punishments would have to be meted out case-by-case with no fixed limits, with the result that both legality and ordinal proportionality would be thrown to the winds. Sexual predators eager for the opportunities of prison life could be punished less severely than freedom-loving pickpockets.

Perhaps, then, the victim's suffering should be dropped from the equation and punishment simply fitted to the wickedness of the wrongful act. But this will not do either, for wickedness is a compound made up of the heinousness of the conduct and the cruelty of the inward disposition, and who can say whether a careless killing of many is more or less wicked than the malicious killing of one?[40] Further, inward dispositions are hidden from us, and crime types as well as the degrees of imputability called intention, recklessness, and negligence are at best crude *indicia* for them. The factors one must consider in judging evil are too numerous and variegated to be captured in a few legal categories, and everyone will weigh them differently.[41] Finally, even if there could be agreement on how wicked a wrongdoer is on a numerical scale, there is no principle for determining how much suffering any particular degree of wickedness deserves; and so convicts will be subject, not to a punishment they have imposed on themselves, but to the arbitrary opinions of the legislator and judge.

Accordingly, we are driven, it seems, to a rule of ordinal proportionality. Let more harmful crimes be punished more severely than less harmful ones; and let intentional inflictions of a particular type of harm be punished more severely than unintentional inflictions of the same type. But this will not work either, for we cannot proceed ordinally without some baseline noncomparative equality, and no such benchmark exists. Moreover, ordinal proportionality is obviously consistent with very cruel punishments: a scale that began with wrist chopping for minor theft and ended with drawing and quartering for murder would be ordinally correct. Thus, no sense can be made of the idea that the punishment must fit the crime.[42]

A defender of legal retributivism might respond as follows. The philosophic truth buried in the saying that "the punishment must fit the crime" is that punishment ought to be nothing other than the criminal wrongdoer's right-denying principle turned against him. Having practically denied an agent's right against interference with its freedom of choice, the criminal cannot complain when his or her own freedom of choice is interfered with in turn. However, this equality between crime and punishment is a purely conceptual one. It is an equality abstracted from any particular kind of crime or any particular measure of punishment. So the question becomes: how can this conceptual identity of crime and punishment be translated into a practical scheme of determinate punishments for determinate crimes?

[40] Hart, *Punishment and Responsibility,* 162.
[41] Ibid, 162–3.
[42] Landau, 'The Failure of Retributivism,' 836–7.

Now, when we seek to translate a conceptual identity into a physical world with qualitative and quantitative dimensions, we cannot expect a perfect replication. For what is identical in the concept may be incommensurable in the world. So we must be satisfied with the best approximation. One possibility is the biblical *lex talionis,* which has always possessed a seductive appeal for retributivists (Kant, for example), because it appears not to be an approximation at all but rather a perfect copy of the retributive idea in the sensuous medium of the world. The metaphysical identity of crime and punishment is rendered in the empirical world as an in-kind identity of physical injuries. But the sensuous replication is actually the crudest just because it is the most sensuous and therefore the least connected to the idea it is supposed to replicate. What is exacted in idea, after all, is liberty for liberty, not physical injury for physical injury. Moreover, because it seeks equality in sensuous terms, the *lex talionis* inevitably leads to absurdities when empirical circumstances bedevil the search for physical exactness (suppose the tooth-basher has no teeth) or when exactness would be shameful for the punisher (suppose the crime was rape).[43]

A modification of *lex talionis,* favoured by Kant, is an equality of suffering that takes into account the social and economic differences of the parties. Thus, argues Kant, a rich man who slanders another must be punished with a humiliation that compensates for his ability to tolerate a monetary penalty.[44] However, if the wrongdoer's pain threshold must be considered, why not the victim's? Must we not punish the thief more or less depending on the wealth of his victim, or the assailant more or less depending on his victim's emotional resilience? But this is to subject the criminal to a measure of punishment determined, not impersonally by public reason, but by the situation or sensibilities of the victim (an analogue to the expensive taste problem in securing equality of welfare).

The best approximation of real-world punishment to the idea of punishment as retribution would be attained if the thing to be equalized in the world corresponded with what is equal in the idea. In the idea, crime is an implicit denial of rights to liberty that entails the criminal's own vulnerability to coercion. Thus, whatever their qualitative and quantitative differences, all material crimes to which retributive punishment paradigmatically applies are interferences with liberty for which the fitting punishment is an interference with liberty.[45] Even a crime against property is ultimately an interference with liberty, since it transgresses the rightful boundary within which the agent's choices hold exclusive sway. Accordingly, deprivation of liberty (which may or may not involve incarceration) rather than corporal punishment or the infliction of suffering for

[43] In this case Kant, ever looking for physical analogues, recommends castration; see *Metaphysics of Morals,* 169.

[44] Ibid, 141–2.

[45] I say "paradigmatically" because treason and interferences with the administration of justice are secondarily subject to retributive punishment as involving challenges to the legal authority making rights possible.

suffering is the mode of punishment best suited to the retributive idea. It is the most civilized of punishments just because it leaves behind the search for empirical equalities of outward injuries or of inward suffering and replicates the *idea* of punishment as retribution for a right denial.

Deprivations of liberty can be distinguished by the length of term during which the convict's liberty is restricted and by the scope allowed for liberty during the term. On their side, crimes can also be distinguished according to the importance to liberty of the interest harmed.[46] Thus, property crimes are less serious than crimes against the body, which are less serious than homicides. The question is whether there is any method for equating a particular (quantitative) term of detention with a particular (qualitative) impairment of liberty by the criminal. And the answer seems to be that, with a few exceptions, there is not. Murder is an exception because it is the only crime for which there is a known and determinate equal punishment. Nothing is equal only to nothing, and so the only punishment equal to a chosen extinguishment of liberty is an extinguishment of liberty.[47] Manslaughter is an exception because the part is less than the whole, and so (depending on the circumstances) an accidental taking of the rest of someone's natural life resulting from a chosen transgression can deserve a taking of liberty for the rest of the killer's life, but a partial impairment of liberty cannot. For the same reason, criminal conduct resulting in a significant permanent impairment of liberty can deserve a life term. All other crimes, however, involve more or less serious qualitative impairments of liberty, and we cannot know what quantitative term of detention (or even range of terms) equals what partial impairment.

Accordingly, the best approximation to the idea of retributive punishment attainable in the real world is an ordinal proportionality ensuring that crimes injuring or endangering interests more important to liberty than those injured or endangered by other crimes are punished with higher maximum restrictions on liberty than those lesser crimes.[48] Culpable homicides on the whole deserve stiffer sentences than culpable bodily injuries, which deserve stiffer sentences than culpable damage to property. Aggravated assault (endangering life) deserves more punishment than assault causing bodily harm. Moreover, intervals between harm types can be rendered ordinally proportionate to intervals between maximum sentences. So, the interval between the maximum punishment for assault and that for assault endangering life should be greater than the interval between the maxima for theft under $1,000 and theft over, and so on.

[46] Of course, the admission of harm at this point stands in tension with the formalist exclusion of welfare from public reason; I discuss this point in the next section.

[47] Yet the very absoluteness of the extinguishment in the case of the death penalty is a reason for doubting the state's right to execute because there is no corresponding absoluteness of knowledge as to whether the defendant is guilty.

[48] On ordinal and cardinal proportionality see Andrew von Hirsch and Andrew Ashworth, *Proportionate Sentencing* (Oxford: Oxford University Press, 2005), 137–43.

Granted, without a benchmark equation at the bottom of the scale, ordinally proportionate punishments could seem cardinally incommensurate with the crimes they punish. One day for theft and two for assault satisfies ordinal proportionality, as does twenty-five years for theft and thirty for assault. The response, however, is twofold. First, our ability to speak of cardinal incommensurability suggests that, even though no principle tells us what quantitative term of punishment a partial impairment of liberty deserves, a cultural sense of fittingness lets us know which punishments are beyond the pale or "unusual". This sense will thus determine the overall level of severity of the ordinal scale. Second, when the idea of retributive justice ceases to guide us, considerations extraneous to that idea can be introduced as a supplement provided their operation is consistent with the idea. Thus, within the bounds of ordinal proportionality, crime prevention concerns can militate against punishments that are too light, while moral concerns about cruelty can prevent punishments that are too harsh.

7. The Limits of Formalism

The formal agency paradigm of penal justice generates a conception of punishment that is compatible with freedom in two senses. First, liability to judicial coercion is implicitly self-willed by the recipient as the logical implication of his or her choice to interfere with a human being's capacity to choose its ends. Having implicitly denied the existence of rights against interference, criminals cannot justly complain if their own rights are violated. Second, the actual imposition of punishment is endorsed by the thinking Agent on behalf of the recipient as the vindication of the practical authority of agency rights against the criminal wrongdoer's challenge to that authority. Inasmuch as retributive punishment observes legal constraints (discussed in following chapters) guaranteeing its self-imposability, it respects the freedom and dignity of the recipient even as it turns the latter's right denial against him. In this way, punishment is distinguished from the violence of the wrongdoer.

Nevertheless, the formalist paradigm lacks the theoretical resources to produce a full account of just punishment. This is so for the following reasons. First, legal retributivism cannot operationalize the conceptual equality of crime and punishment in a practicable scheme of determinate punishments without tacitly acknowledging what its framework officially denies: the existence of material interests objectively important to freedom (hence to all agents) and so the existence of a public conception of welfare and of legally cognizable harms. If one were to adhere strictly to the formalist equation of the public interest with liberty, one would have to say that there is only one crime—an intentional interference with liberty implying a denial of rights to liberty—and that all crimes, however qualitatively and quantitatively differentiated according to the type and extent of the harm inflicted or attempted, must be punished identically. Of course, no

one—not even the staunchest formalist—would seriously defend such a view. Even those who deny the relevance to penal desert of harmful outcomes because they are attributable to chance rather than to choice would vary punishments according to the harm intended or attempted.[49] But the inevitable grading of crimes according to the "seriousness" of the harm inflicted stands in tension with the formalist equation of material interests with subjective interests; hence it foreshadows a framework of penal justice ordered to a new conception of public reason, one that takes explicit account of the commonality to agents of some interests and harms.

Second, while the formalist paradigm generates a retributive idea of punishment requiring that punishability be implicitly self-willed through a chosen interference with choice and that the punishment mirror the crime, it generates no requirement that "the crime" be defined only by material outcomes that are themselves connected to the wrongdoer's agency. So, once liability to punishment *simpliciter* is self-willed by agents by virtue of a chosen interference, they become liable to state coercion directed at crime control for all resulting proscribed harms no matter how remote and unforeseeable. We see this, for example, in the thin-skull rule and in the definition of manslaughter as requiring the foreseeability only of nongrievous bodily harm.[50] In both cases, the crime description and measure of punishment are unconstrained by a requirement that the consequence for which the agent is punished be imputable to its agency in some degree, whether by intention, foresight, or foreseeability. Of course, these doctrines serve deterrence and the disabling of human threats; but the only reason these ends have full play here is that there is no requirement that punishable consequences be imputable to agency to limit their operation. Why should this be so? Why should it be necessary that punishment as such be connected to the wrongdoer's agency but not the outcomes for which he is punished?

The answer lies in the formalism of the formal agency paradigm. Given formalism's equation of the public interest with freedom of choice, it protects the agent against interferences with its liberty, but it does not protect the agent against vulnerability to blind chance—in this case to causal chains initiated by its culpable wrong that are beyond its power to foresee and control. Such protection is possible only under a richer conception of freedom than the one informing the formalist framework, one that understands by freedom not only undetermined choice but also self-determination, or the power to shape one's life in accordance with self-authored ends. Accordingly, once the agent has invited punishment through a chosen interference with choice, there is no public reason to stop its liability for all chance consequences of the wrong for the sake of deterrence, incapacitation, or whatever.

[49] See AJ Ashworth, 'Sharpening the Subjective Element in Criminal Liability,' in RA Duff and NE Simmonds, eds., *Philosophy and the Criminal Law* (Wiesbaden: Cambridge University Press, 1984), 79–89.

[50] *R v Holland* (1841) 2 Mood & R 351; *R v Larkin* (1942) 29 Cr App R 18, 23.

Third, the formalist paradigm generates no justification for a right infringement other than self-defence (defence of property, defence of others) and punishment for a crime. A right of self-defence is certainly intelligible within the formalist framework, since a right against interference with liberty is a right to defend oneself or others against transgressions if the public authority cannot intervene and provided one uses no more force than is necessary to repel the invader (see Chapter 6). However, justifications based on a comparison between the harm inflicted on an innocent and the harm avoided to a wrongdoer (for example, necessity) have no place in a theoretical framework for which all ends are subjective, all harms are relative to preferences, and for which liberty is the only public thing. If A values B's loaf of bread more than B does, then a bargain is in order, not a taking. The fact that A values B's bread so highly because he is starving and has a powerful aversion to death is here of no more consequence than if he desired the bread to satisfy a craving for carbohydrates. Since all ends are subjective, they are on a par. Accordingly, justifications other than self-defence presuppose a framework of penal justice ordered to a conception of freedom as self-determination—one for which objective harms are possible as harms to interests essential to acting from self-authored ends.

Finally, the formalist paradigm is powerless to regulate liability for offences (called public welfare offences) not involving either a right infringement or a violation of laws (for example, against treason, bribery of officials) ministerial to rights protection. While formalism insists that a public policy of facilitating the agent's pursuit of satisfactions be subordinate to rights protection, it generates no constraints on liability for breaches of statutes having the general happiness as their aim. This is so because formalism's retributivist legitimation of penal force as implicitly licensed by the recipient has no application to the actions and omissions proscribed by these statutes. Since there is no right infringement in selling alcohol to a minor, for example, there is no invitation to a retaliatory right infringement in choosing to do so, nor can the point of penalizing be to vindicate the reality of rights. In this context, penal force is justified by prevention rather than by desert.

However, if public reason is identified with respect for liberty and if legal retributivism is thus considered an exhaustive account of penal justice, then the penalization of public welfare offences will occur in a normative vacuum. No liability-constraining rule will supplement the retributivist one for crimes, and so penalizing the faultless will be left free for rationalization in cost–benefit terms. Put otherwise, if there is no logical circuit between the intentional doing of the proscribed action and the coercibility of the doer, then no one can either deserve or not deserve punishment for these actions. Thus penal force cannot here be constrained by desert. But if penal justice is understood solely in terms of this circuit, then the penalization of activity not involving a right infringement will be constrained by nothing except cost–benefit rationality. And that rationality will likely favour no-fault liability, since the penalty and stigma for conviction

are often light, while deterrence is perhaps much enhanced by the greater ease of conviction. Thus, the dichotomy in the penal law of many jurisdictions between a requirement of wilfulness for crimes and no-fault liability for public welfare offences is an offshoot of the formalist paradigm taking itself to be exhaustive of penal justice.

Still, these limitations of the formalist paradigm are not a reason for jettisoning it. They point only to the insufficiency of the framework as a full account of penal justice and hence to the need for integrating it within a more comprehensive account. That formalism is compelled to recognize tacitly the objectivity of some goods it officially denies is a reason for demoting it from an exhaustive theory of penal justice to a constituent one; it is not a reason for dispensing with it altogether. Moreover, to dispense with it would be to lose its contributions to a practice of punishment consistent with freedom, for we will see that the framework that supplies what formalism lacks cannot provide on its own what formalism contributes: an immovable desert-based constraint on punishment for crimes. It is thus a grave mistake to think that the valuable elements of the formalist paradigm can be salvaged within a new framework while the paradigm itself is sunk. How the formalist paradigm can be coherently preserved within a more inclusive framework remains to be seen.