LEX TALIONIS

Jeremy Waldron*

The approach to punishment known as *lex talionis* (LT) — "an eye for an eye, a tooth for a tooth" — is (fortunately) not as silly as it sounds. Though the principle retains its attraction for defenders of capital punishment ("a life for a life"), people think they can discredit it almost immediately by asking "What penalty is to be imposed on the *rapist*, according to this principle?" Amidst the general hilarity that follows, the speaker is able to put LT quietly to one side and move on, as he thinks, to some more plausible version of retributivism.[1]

In this Essay, I shall argue five things. First, since LT is a principle about what counts as an appropriate punishment, it is compatible with a variety of theories about the point or justification of punishment, including utilitarian theories. In Section I, I shall sketch out a couple of these. Second, LT cannot be thought to require that *the very same action* that constituted the offense should be visited as punishment upon the offender. Rather, the requirement must be that the act of punishment be *similar* to the offense in certain respects. Which respects these should be is a matter of normative argument.

On the basis of those two general claims, I shall argue for the following propositions: Third, a defense of LT (even for murderers) is consistent with a rejection of capital punishment in cases of homicide. Fourth, the operation of the principle need not be confined to those cases where direct harm, suffering or loss has been caused to a particular assignable individual. It can be applied also to offenses like perjury, tax evasion and blackmail. Fifth, there is no incompatibility between the stern dictates of LT and the more humane doctrine that punishment should be adjusted to reflect the degree of the offender's responsibility for his crime.

Now the third, fourth and fifth positions look quite implausible, if only because they contradict our familiar image of the advantages and disadvan-

---

* Professor of Law, Jurisprudence and Social Policy Program, School of Law (Boalt Hall), University of California, Berkeley. B.A. 1974, LL.B. 1978, University of Otago, New Zealand; M.A. 1980, D. Phil. 1986, University of Oxford.

The main argument in Section II of this paper developed out of a conversation with Heidi Hurd, Bruce Mann, Michael Moore and Elizabeth Warren. I am grateful to them for stimulating the argument, and also to Henry Shue for helping me develop it. The bulk of the paper was drafted under the auspices of the Program on Ethics and Public Life at Cornell University; I am grateful to the Program and its Director for support and assistance.

1. For example, this is the approach taken in Hugo A. Bedau, *Retributivism and the Theory of Punishment*, 75 J. PHIL. 601; 611 (1978). HUGO A. BEDAU, DEATH IS DIFFERENT 262-63 n.61 (1987) [hereinafter DEATH IS DIFFERENT], provides a list of recent retributivist theories and claims that all but a few of them reject *lex talionis* out of hand.

tages of LT. However, I shall show that once we accept the second of my general claims — about what LT can and cannot be thought to require — we will see that claims III and IV fall naturally into place. And I shall argue that claim V only looks implausible if we divorce LT from any consideration of what might be said in its favor. If we base our interpretation of the principle on the considerations that might justify its adoption, then we will see that those considerations leave plenty of room for the more conventional and familiar aspects of our penal law and practice.

I

What is *lex talionis*? It is a theory that purports to guide us in our choice of appropriate penalties. In our civilization, it is familiar from Mosaic formulations such as, "An eye for an eye, a tooth for a tooth," and "[Y]e shall take no satisfaction for the life of a murderer, which is guilty of death: but he shall be surely put to death."[2] But LT is not just a principle about eyes, teeth and murder. What these formulae have in common is the idea of *doing to the offender, as punishment, what the offender did to his victim.* There are two actions to consider: the act of offending and the act of punishment. LT holds that they should be the same; and since the act of punishment is the only one over which we have any control, we should choose a punishment that matches the character of the offense. As a general principle, we might try formulating it as follows:

> The action which constitutes punishment for any offense should be the same as the act which constituted the offense.[3]

It should be the same act, but of course this time the offender should play some role other than that of the agent. This time, preferably, he should be one of the victims of the act which constituted his offense.

In the popular imagination, LT is regarded as a principle of retribution. Its stern insistence on matching the penalty to the crime seems to indicate an almost entirely backward-looking approach to punishment. We gouge out the offender's eye because he gouged out the victim's eye. We take his life because he took someone else's life. The fact that the reason for punishment is given in the past tense seems to indicate that this is not a consequentialist account. Looking forward to the likely consequences of inflicting a penalty, we might be tempted to modify it, to make it more severe for the sake of

2. The passage about murder is from *Numbers* 35:31. *See also Exodus* 21:23-25 ("And if any mischief follow, then thou shalt give life for life, eye for eye, tooth for tooth, hand for hand, foot for foot, burning for burning, wound for wound, stripe for stripe."); *Leviticus* 24:17-20 ("And he that killeth any man shall surely be put to death.... And if a man cause a blemish in his neighbour; as he hath done, so shall it be done to him. Breach for breach, eye for eye, tooth for tooth...."). But the Mosaic codes are far from consistent in all this. Sometimes they exact more draconian penalties than *lex talionis* would indicate, for example, "He that smiteth his father or his mother, shall be surely put to death" (*Exodus* 21:15) and "If a man shall steal an ox, or a sheep, and kill it, or sell it; he shall restore five oxen for an ox, and four sheep for a sheep." *Exodus* 22:1. Versions of the principle are also found in the Code of Hammurabi; for a discussion, see MARVIN C. HENBERG, RETRIBUTION: EVIL FOR EVIL IN ETHICS, LAW, AND LITERATURE 62-68 (1990).

3. We shall modify this definition somewhat in Section II, where I address the ambiguities in the phrase "same act."

deterrence, or less severe when severity can do less good. But LT seems to repudiate such considerations in its insistence on a simple match between penalty and past offense. It therefore seems incompatible with utilitarian and other consequentialist approaches to social decisionmaking.

That's how things seem. Three points, however, should make us pause before peremptorily rejecting the idea of a consequentialist defense of LT. First, there is the obvious point that acts of punishment that fulfill LT cry out for justification. If the original offense was wrong and unjustified, why isn't its repetition wrong and unjustified? The principle does not by itself tell us the answer to this question, but clearly it requires an answer. Secondly, there is a point that H.L.A. Hart has stressed in his work on punishment: you cannot infer anything about the General Justifying Aim of punishment from looking at a principle governing the amount of punishment to be inflicted. Theories of punishment are complex and articulated, and it is possible for the background aim to be utilitarian while the rules governing distribution and amount appear more retributivist in character.[4] Thirdly, we know from modern utilitarian theory that it is possible for deontological-sounding rules to be supported by consequentialist considerations. That is what goes on with indirect utilitarianism.[5] Before we rush to the conclusion that LT is necessarily deontological in its basis as well as its application, we should consider what might be said for it in an indirect utilitarian approach to social decisionmaking.

The first point, about the need for a justification, has two sides to it: there is the general problem of justifying punishment, and there is the particular problem of justifying this principle, as opposed to some other principle, for determining the type of penalty to be imposed.

The general problem is particularly acute in the case of LT. A justification of punishment is required in all cases because punishments usually have features that, in general, make actions impermissible. That an action is the deliberate infliction of harm or suffering on one human being by another, or a deliberate restriction of his liberty, or the deliberate taking of his life or property, is usually a reason for judging it wrong. Justifications of punishment must show some special reason why that judgment does not hold in a penal context. That is true of punishment generally, and it is *acutely* true of punishment understood in terms of LT. For if the act of punishment is precisely the same act as the one condemned as an offense, then the very terms of its imposition (that it is imposed *for* an offense in virtue of the wrongness of the offense) seem to *entail* the wrongness — indeed the punishability — of the act of punishment itself, unless more is said.

Retributivists sometimes claim that they can avoid this onus of justification, since — unlike their utilitarian opponents — they are not committed to the general proposition that harm, loss and suffering are evils that always count against the actions that inflict them.[6] Retributivists say they are not

4. H.L.A. HART, PUNISHMENT AND RESPONSIBILITY: ESSAYS IN THE PHILOSOPHY OF LAW 8-11 (1968).

5. Most notably in the work of R.M. Hare: for example, RICHARD M. HARE, MORAL THINKING: ITS METHOD, LEVELS AND POINT (1986).

6. The most famous example is GEORG WILHELM FRIEDRICH HEGEL, PHILOSOPHY OF RIGHT ¶ 99 (T.M. Knox, trans., 1967) (remark).

against the infliction of harm or suffering as such. They are against it only when it happens to the innocent, not when it happens (in due proportion) to the guilty. Similarly, a defender of LT might say that he is morally opposed only to the taking of innocent life, not to the taking of life generally, or only to the knocking out of the teeth of those who have not knocked anyone else's teeth out, not to all dental injuries.

But this reasoning is fallacious. The idea of punishment is logically posterior to a theory which defines wrongdoing, just as the idea of desert is logically posterior to a theory which defines deserving and undeserving conduct or character. The ideas of guilt, innocence and punishment cannot feature in the first principles of such a theory, though, of course, once the first principles are in place and the idea of punishment has been justified, then they may be used in any statement of moral principle that sums up the gist of the whole. For example, in order to establish a justification of capital punishment for murder, we must (1) explain why it is wrong for one human being deliberately to take the life of another, and (2) explain why taking the life of a murderer as punishment is not wrong notwithstanding (1). We may sum up the effect of (1) and (2) by saying (3) that what is wrong is for one human being to deliberately take the life of an *innocent* person, but we cannot appeal to (3) in discharging the burden of justification laid down in (2). The concepts of guilt and innocence involved in (3) cannot be relied on unless the tasks laid down in (1) and (2) have already been performed.

LT also needs justification against principles that accept the idea of punishment but determine the appropriate penalty in other ways. It needs justification, for example, against theories of moral desert — which try to ensure that the distribution of suffering and happiness in the community is proportional (in some quantitative sense) to vice and virtue. It needs justification against theories which see punishment as a way of readjusting the equal distribution of benefits and burdens in society. And of course it needs justification against utilitarian theories of deterrence and rehabilitation which base the amount and type of punishment on consequential considerations.

To discharge both these burdens, the defender of LT must try to explain the point of his otherwise irrational obsession with matching the act of punishment to the act which constituted the offense. Explaining the *point* of an action is not necessarily the same as justifying it in terms of its consequences. An action may look silly, cramped, and unnecessary until it is characterized as the act of putting in golf. Described as a move in a game, it is seen in a different light, and the original challenges to it now seem beside the point. But that is not a consequentialist defense. Similarly, in the moral realm: an act that might seem wasteful and even corrupting (giving twenty dollars to a derelict who I know will buy alcohol rather than to a homeless parent who will buy food for her children) is cast in a different light when I explain that it is money owed to the derelict from a transaction a few years back. Again, the justification is in terms of the moral character of the act, not in terms of consequences.

The best example of a non-consequentialist account of the point of LT is provided by Immanuel Kant.[7]

> What kind and what degree of punishment does public legal justice adopt as its principle and standard? None other than the principle of equality (illustrated by the pointer on the scales of justice), that is, the principle of not treating one side more favorably than the other. Accordingly, any undeserved evil that you inflict on someone else among the people is one that you do to yourself. If you vilify him, you vilify yourself; if you steal from him, you steal from yourself; if you kill him, you kill yourself....
>
> But what is meant by the statement: "If you steal from him, you steal from yourself"? Inasmuch as someone steals, he makes the ownership of everyone else insecure, and hence he robs himself (in accordance with the Law of retribution) of the security of any possible ownership.[8]

The idea seems to be that punishment involves an exercise in universalization. In doing to the offender what he did to others, we show what it would be like for him if everyone did that.

Putting the action in this light gives it some plausibility: certainly the best defenses of LT are going to involve reciprocity and universalizability at some point. But it is less easy to see why Kant describes punishment along these lines as a *categorical imperative*, and why he insists, in a famous passage, that:

> Even if a civil society were to dissolve itself by common agreement of all its members (for example, if the people inhabiting an island decided to separate and disperse themselves around the world), the last murderer remaining in prison must first be executed....[9]

He goes on to give two reasons: (i) "so that everyone will duly receive what his actions are worth" and (ii) "so that the bloodguilt thereof will not be fixed on the people."[10] Neither is helpful: (ii) is a reason only if we already have an account of why the punishment must be inflicted, and (i) simply begs the original question of *why* everyone must receive what his actions are worth (in this or any other sense). Kant considers an explanation later taken up by

---

7. To say that an account is the best available is not necessarily to say that it is a good account. For some (justified) pessimism about the prospects for a charitable reconstruction of Kant's theory, see Jeffrie G. Murphy, *Does Kant Have a Theory of Punishment?*, 87 COLUM. L. REV. 509 (1987).

8. IMMANUEL KANT, THE METAPHYSICAL ELEMENTS OF JUSTICE 101-02 (John Ladd trans., 1965).

9. *Id.* at 102.

10. *Id.* The idea of blood-guilt seems to be biblical: "So ye shall not pollute the land wherein ye are: for blood it defileth the land: and the land cannot be cleansed of the blood that is shed therein, but by the blood of him that shed it." *Numbers* 35:33.

Hegel[11] — that the offender wills his punishment, either by committing the original offense or in signing up for the social contract, but he sensibly rejects it: "No one suffers punishment because he has willed the punishment, but because he has willed a punishable action."[12]

The difficulty for Kant is that there is no general requirement (in his system or anyone else's) that we should always do what we can to universalize other people's actions, so that they experience them from another point of view; therefore, Kant must show that there is something special about wrongdoing that requires this. In *The Metaphysical Elements of Justice*, he argues that an offender has no right to complain if we do act towards him as he has acted towards others:

> [N]o one has ever heard of anyone condemned to death on account of murder who complained that he was getting too much [punishment] and therefore was being treated unjustly; everyone would laugh in his face if he were to make such a statement.[13]

But this is not the same as showing that we have a duty to act towards him in this way.

One explanation of the point of this sort of punishment-as-universalization is as an exercise in moral education. Sometimes punishment is justified on the ground of its communicating a message to the offender: "This is how wrong what you did was." Indeed, in the case of punishment governed by LT, we might want to communicate the message: "[T]his is the wrong action that you did."[14] If someone asks why these messages cannot be conveyed by letter or over the telephone rather than through punishment, the response may be that the offender has already shown by his conduct that he has not internalized the message in any of its conventionally communicated forms, and we now have to abandon words and try to show him why such acts are wrong, in the most direct and powerful way, by getting him to experience them at the sharp end, so to speak.[15]

It is still hard to see why this would be a *categorical imperative* (as opposed to *a rather good idea*). But if we move outside the Kantian framework, we can see this educative approach as a sensible thought that might lie behind LT. It is hard to know whether to classify it as consequentialist. Certainly, it takes moral education and reform as a forward-looking aim, and so it is consequentialist in that sense. But the moral principles being taught

---

11. *See* HEGEL, *supra* note 6, ¶ 100: "The injury [the penalty] which falls on the criminal ... is *eo ipso* his implicit will...." And in the "Addition" to that paragraph: "[T]he criminal gives his consent already by his very act."

12. KANT, *supra* note 8, at 105.

13. *Id.* at 104. At the end of the work, he writes:

> "The only time a criminal cannot complain that he is treated unjustly is when he draws the evil deed back onto himself [as a punishment] and when he suffers that which according to the spirit of the penal law — even if not to the letter thereof — is the same as what he has inflicted on others."

Id. at 133.

14. ROBERT NOZICK, PHILOSOPHICAL EXPLANATIONS 370-71 (1981).

15. For an excellent discussion of the communicative theory of punishment, see Herbert Morris, *A Paternalistic Theory of Punishment*, 18 AM. PHIL. Q. 263 (1981).

may be deontological in character (and this is likely to be so if universalization is the crux of the lesson).

In his book, *Well-Being*, James Griffin has argued that any normative theory — utilitarianism included — has to have a place for notions of guilt and repentance, that is, for an agent's negative response to (what the theory regards as) his own wrongdoing.

> Punishment has a rich life of its own, quite independent of social institutions. It starts with the appropriateness of a person's response to his own wrongdoing. This, since it is purely a one-person affair, is not yet what can be regarded literally as "punishment." The appropriate response to wrongdoing is: perception of wrongdoing, guilt, and repentance. But it influences the appropriateness of interpersonal responses to wrongdoing, which certainly can be called "punishment." That is true in the one-person case; it is also true in the interpersonal case. My response to your wrongdoing is appropriate only when, and to the extent that, it contributes to your going through the same process: perception, guilt and repentance.[16]

Although social punishment for a utilitarian may also take on other functions (of deterrence, prevention and reform), these notions of perception, guilt and repentance remain at its core, and the institution is always in tension when the additional tasks pull the process too much away from what is appropriate to those ideas. (Notice, however, that the various tasks are not unconnected: reform is linked with perception and repentance, and prevention and deterrence with the negative character of perception and guilt.)

Within this framework, one can see how LT might be regarded as a good idea even within a utilitarian theory (though of course utilitarians are unlikely to commit themselves comprehensively to LT, to the exclusion of other penal strategies). Human nature may be such that the vivid lesson conveyed by "turning the tables" on the offender is the quickest route to awareness of wrong-doing and repentance. Utilitarianism requires of its agents sensitivity to the suffering of others, and indeed models its decision-making on the idea of a super-person who experiences both the advantages accruing to the agent from the offence and the pain accruing to the victim.[17] For these reasons, it may be thought important for the wrongdoer — who has otherwise proved impervious to moral instruction — to *experience* the suffering his act has inflicted on others. Thus, reasoning along the lines of LT might have a part to play in the determination of an appropriate penalty even for a utilitarian judge or legislator.

I don't want to develop these justifications any further in the present Essay. It is not my intention to argue that LT is the best principle for choosing punishments, only that there is more to be said for it than is commonly

---

16. JAMES GRIFFIN, WELL-BEING: ITS MEANING, MEASUREMENT, AND MORAL IMPORTANCE 272 (1986).

17. See the discussion of the "ideal observer" model in JOHN RAWLS, A THEORY OF JUSTICE 22-27, 183-92 (1971). *See also* RICHARD M. HARE, FREEDOM AND REASON 112-36 (1963).

supposed. Awareness of these justifications will be important, however, as we move on to discuss exactly what the principle requires.

II

Lex talionis requires that the act of punishment be the same as the act that constituted the offense. But is it possible for the same act to be performed twice, at different times and with different *dramatis personae*?

It depends whether one is talking about particular actions or act-types.[18] Consider the particular action that is my writing the previous sentence for the first time: on February 18, 1990 at 3:07 p.m., in Ithaca, New York, Waldron's fingers moved over the keyboard of his computer in the pattern "Shift-I, t, space, d, e, p, ... etc.." That action cannot be performed again, by me or anyone else. It is a particular event that happened at a time and place, with particular persons, equipment, and movements. But an act *of the same type* may be performed again: someone else may key exactly the same input into her computer a week later. Her particular act will then share certain general features with mine. My action in writing that sentence was an instance of various kinds — typing, writing part of an article, using a computer, wasting everyone's time, using WordPerfect 5.0 — and other particular actions may be of all or some or none of those kinds.

So the phrase "the same act" is ambiguous with regard to this incident. It *may* mean the same particular action, as when we ask, "Was Waldron's *thinking* about the formulation of that sentence part of the *same act*?" But if we know that a different agent is being contemplated, then "the same act" *can only mean* a different particular action of the same type (as in, "Will *the same act* have to be performed several times by different people in preparing this Essay for publication?"). And then the obvious question will be: "What similarities — that is, which of the various types or kinds exhibited by the original action — do you have in mind?"

LT is the principle that the action visited as punishment upon a criminal should be the same as the action that constituted his offense. In almost every case it is evident that the action visited as punishment is to take place at a different time and with different *dramatis personae* from those of the original offense.[19] Punishment characteristically takes place *after* the offense, and the whole point of LT is that the criminal should occupy a role other than that of agent in the re-enactment that his punishment involves. It follows from what I said in the previous paragraph that the principle can only be interpreted as requiring that an action *of the same type* as the offense be visited on the criminal as punishment. It follows also that we will expect the punishment

18. This is not the same as the distinction between act-tokens and act-types in ALVIN I. GOLDMAN, A THEORY OF HUMAN ACTION 10-19 (1970). The notion of act-type is the same, but for Goldman, an act-token is the instantiation of a particular type or property. Thus, in the example given in the text: Waldron's moving his fingers and Waldron's typing "It depends" are separate act-tokens on Goldman's analysis, whereas on my account they are the same particular action (which instantiates many properties). The analysis I have given, then, is closer to that of Donald Davidson, *The Logical Form of Action-Sentences*, *in* ESSAYS ON ACTIONS AND EVENTS 105 (1980).

19. The only exception is the rare case in which the offense is (a) a self-regarding one (like attempted suicide) and (b) appropriately regarded as its own punishment.

demanded by LT to share some of the features of the original offense and not others.

This is not a matter of moving from a literal to a relaxed or metaphorical understanding of the principle. "The same act" is simply ambiguous as between "same particular action" and "same type." Since "same particular action" *does not make sense* with regard to crime and punishment, the phrase can only mean "same type." There may then be more or less relaxed ways of understanding *that*. As we shall see, which similarities are important (and how many) will depend in part on our reasons for deploying the principle.

I emphasize that point because it is important to see that the proponent of LT is not committed to the idea that there is a fundamental and irreducible similarity-relation between pairs of acts in terms of which we can say, "This act definitely is (or is not) *the same as* that." A given pair of particular actions may exhibit several similarities and several dissimilarities, and all that means is that the members of the pair share certain features and not others.[20] That is all there is to be said. There is no sense to the further claim, "Therefore, on the whole, they are similar (or dissimilar)."

What features does the proponent of LT have in mind when he says the punishment should be of the same type as the offense? In detail, the answer will depend, as I have indicated, on our reasons for adopting the principle. As we have seen, one prominent line of argument for LT involves Kantian universalization. Respect for the criminal's agency is thought to require us to act toward him in light of the universalization of what we take to be the maxim of his action in committing the crime. But of course universalization always involves selecting some relevant feature(s) of the act in order to specify the maxim that is to be universalized. The process requires a choice among the features of the universalized act; it involves focusing on some and ignoring others. Something similar is true of communicative theories. If the point of punishment is to teach the criminal a lesson in some vaguely literal sense, then we shall have to decide what it is we want to say to him about his offense as we re-enact it in his punishment. Once again, we have to be selective, so that the message we convey is clear, effective and unencumbered.

Is there anything we can say at a general level about the choice of appropriate features? To begin with, an action which is an offense will have certain *deontic* characteristics, namely, its unlawfulness, its wrongness, and perhaps its being the violation of a right and its not being the exercise of one. Is LT likely to require that the punishment share any of these features? No, but the negative answer must be stated carefully. Since LT is intended to operate as part of a theory of lawful punishment, it cannot require the carrying out of unlawful acts. That is compatible, however, with the act of punishment having characteristics which would be sufficient to mark it as unlawful if it were performed in any other context. In the same way, since the principle is intended to operate as a theory of rightful or morally justified punishment, it is

20. I use "characteristic" and "feature" as general terms to refer to the semantic values of many-place as well as one-place predicates. That an event happened quickly is one of its features; that it involved a betrayal of one person by a second person to a third is another.

evident that it cannot require the carrying out of actions which are wrong.[21] But again, that is compatible with the act of punishment having characteristics that would make it immoral if it were not being performed as punishment.

This suggests a distinction between several different sets of features an offense might have. First there are what I have called its deontic features — its unlawfulness, its wrongness, its not being the exercise of a right, its display of vice rather than virtue, and so on. A second set of features are those on which members of the first set are supervenient — the characteristics that *make* it wrong, impermissible, vicious, etc.. These I shall call its wrong-making features. A third set might comprise those features of the action on which its deontic properties are not supervenient, attributes which are irrelevant from the legal or moral point of view. Call these its irrelevant features.

Thus, consider a particular act of murder. It has wrongness as a deontic feature. It has the feature of being an intentional killing of one human being by another as a wrong-making characteristic. And it may have features such as involving antique weapons or happening on a Tuesday, which are irrelevant from the moral point of view.

What we have established so far is that LT will not require that the act of punishment should share any deontic features with the original offense, for that would play havoc with the wider theory of the justification of punishment of which we expect it to be a part. But it may nevertheless share some features of the second class, the features of which the deontic characteristics are supervenient. The punishment must not be wrong (even though the offense was), but it may still share features with the offense that made the latter wrong. Indeed, on the justifications we considered earlier, that is the whole point: by putting the offender in a role other than that of agent, we show him what was wrong about the offense.

What about the third class of properties — those (such as happening on a Tuesday) that are irrelevant to the moral quality of the original offense? Must these or any of these be reproduced in the appropriate punishment? The answer is surely "No." What point could be served by requiring replication of morally irrelevant features? As we have seen, the best defenses of LT are based on reciprocity or universalizability: the idea is to bring home to the offender (either communicatively or literally) what made his action wrong. Now, to reciprocate or to universalize is always to select features of the act being reciprocated or universalized, and if the aim is to bring the wrongdoing to bear on the offender, it is unlikely that we will be interested in reproducing features that were irrelevant to our concern about the action in the first place. Of course, the act of punishment also will have certain features that are irrelevant to its point or purpose: it too will happen on a certain day of the week, and be administered by a person of a certain hair color, etc.. But there is little danger of those features detracting from the point of the punishment, and such danger as there is would be exacerbated rather than mitigated by any deliberate attempt to match the irrelevant features of the punishment to the irrelevant features of the offense.

---

21. Or at least wrong, all things considered, which is presumably the deontic property it would have to share with the original offense.

Of the three classes of features of an offense that we identified — its wrongness, its wrong-making features, and its morally irrelevant features — it is the second class that LT will focus on. LT can be restated as a theory that requires the act of punishment to possess some or all of the wrong-making features of the original offense:

> The action which constitutes punishment for an offense should possess some or all of the characteristics that made the offense wrong.

Now, actions can be characterized in terms that are general and abstract or in terms that are specific and relatively vivid. This is not the same contrast as the one we have had already, between act-types and particular actions; instead, it is a contrast that discriminates among act-types. A given particular action may be an instance of all of the following types: (i) the killing of one human being by another, (ii) the fatal shooting of one human being by another, and (iii) the fatal shooting of one human being by another with an semi-automatic weapon. The third of these types is more specific than the first, but it is still a type that any number of particular actions might instantiate.[22]

Suppose a given particular satisfies all three of these types. If the action is wrong, it is wrong in virtue of one of these features, but which one? Our inclination is to say that it is wrong in virtue of its being an act of type (i) — the killing of one human being by another. It is the killing that makes it wrong, not the fact that the homicide involved shooting or that it involved the use of an automatic weapon.[23] Now, knowing that the act belonged to either of the other two types, (ii) or (iii), involves knowing that it was a homicide, so we cannot say that those features are morally irrelevant. However, the knowledge that it belonged to type (ii) or type (iii) imparts information about features of the action that are, strictly speaking, irrelevant to its wrongness (*qua* homicide). So we should identify the relevant act-type at the level of generality that indicates exactly the features that make the action wrong.

Thus, suppose someone were to focus on the act-type (ii): that the action was the fatal shooting of one human being by another. We might ask, "What makes an act of that type wrong?" The answer would be, presumably: "It is not the fact that it's a shooting that makes it wrong, it's the aspect of homicide — the killing of one human being by another." This indicates that the wrongness really supervenes upon the property of being a killing, not the property of being a fatal shooting.

It is likely, then, that a proponent of LT will focus on the feature that defines type (i) as the one to reproduce in his act of punishment for this offense, rather than types (ii) and (iii). That is, he will think it important that

---

22. For the distinction between the universal/particular contrast and the general/specific contrast, see HARE, *supra* note 17, at 37-48. The former contrast corresponds to the type/token distinction, but it is the latter contrast (general/specific) that we are discussing in this and the following paragraphs.

23. Of course, discharging a firearm and possessing or using a semi-automatic weapon may also be offenses considered in themselves, apart from the homicidal aspect. To the extent that our interest is confined to those offenses, the wrongness of the act-token will focus on (ii) and/or (iii), ignoring (i).

killers be put to death, rather than that shooters be shot, or that killers who use semi-automatic weapons be executed with firearms of the same type and caliber.[24]

Someone may try to push us to a level of generality that is higher than (i): "What makes an act of killing wrong? What is it about killing that makes it something that one should not do?" Suppose, for example, that I deliberately injure someone in a way that is calculated to deprive them of consciousness and reduce them to the state of a living vegetable, from which medical science is incapable of awaking them. Isn't that as bad as killing, and in roughly the same way? In which case, the bad thing about a killing is not that it involves the infliction of what clinicians define as death, but that it disrupts and terminates the conscious self-direction of one's life. From this point of view, the fact that clinical death is involved may seem as irrelevant as the fact that shooting is involved.

If anyone takes that line, it may alter his views on what is appropriate punishment for murder. The aim will be to perform as punishment an action calculated to disrupt and terminate the conscious self-direction of the prisoner's life. Such punishment may be something other than capital punishment as we understand it: putting the prisoner into an irreversible coma, for example, or certain forms of very highly restrictive confinement. Of course, this new approach does not *preclude* execution (any more than focusing on (i) rather than (ii) precluded the use of firing squads). But it may give us other alternatives for punishing murder, just as states may choose among various forms of execution.

I will return to the issue of capital punishment in Section III. But the points just made help deal with the perennial conundrum of the *lex talionis* approach to offenses like rape. A particular act may belong to the following three types: (i) it is an unwanted sexual violation of one person by another; (ii) it is an act of sexual intercourse by a man with a woman without her consent; (iii) it is an act of sexual intercourse by a man with his girlfriend without her consent.

Each of these types indicates morally relevant features, though we may well think that (iii) indicates also a feature that is irrelevant from the moral point of view. Type (ii) is less specific than (iii), and it is usually taken as the feature on which the unlawfulness of rape supervenes. But, obviously, there are problems for LT in regarding (ii) as the crucial feature. If it is essential to the wrongness of rape that the offender be a man and the victim a woman, then it will always be literally impossible to rape the rapist. Those who are interested in persevering with LT are likely therefore to want to move from

---

24. But one has to be careful here. Moving to a higher level of generality — e.g., from "fatal shooting" to "homicide" — is not always desirable in one's penal thinking. Sometimes it is important to hang on to some of the specifics of the offense, in order to get an appropriate sense of its wrongness. For example, though "robbery" is a more abstract category than "robbery with violence," still the latter characterization better captures our sense of the wrongness of certain crimes. So my suggestion is not that the more abstract characterization is always to be preferred, but that we should position ourselves at whatever level of abstraction best captures our conviction about the wrongness involved.

(ii) to the more abstract (i), which is not artificially restrictive of gender in this way.

But that still leaves us with the unsavory requirement that the act of punishment be a sexual violation (of the offender). It is useful, therefore, to ask whether it is possible to obtain an account of the wrong-making characteristic which is even more abstract, even less specific, than (i). Some feminist writers have argued that rape should be reconceptualized as a crime of violence, not as a specifically sexual offense.[25] If such an argument succeeds, it will establish that the specifically sexual aspect of rape is as irrelevant to the violence and hence the wrongness of the offense as the fact that, in other cases of assault, a blow is struck with a fist rather than a boot. Once this is established, it may be possible to omit the sexual element altogether as the essence of just punishment for rape according to LT.[26]

In the end, I think it unlikely that an adequate characterization of the wrongness of rape can be given purely in terms of assault. It may be essential to our understanding of the offense that it involve trauma, indignity and degradation. But again these elements of the offense can be understood in more or less abstract terms. The level of abstraction we choose will be our sense of what ultimately matters in our reckoning something an offense, our sense of finally why it is wrong. Since the proponent of LT aims to base his choice of punishment on the best available understanding of the wrongness of the offense, he will necessarily want to follow the process of this abstraction, step by step, as far as it can plausibly be taken.

I hope it is clear where this line of argument is heading. We have reached two conclusions. First, LT requires the act of punishment to exhibit some or all of the wrong-making features of the offense, some or all of the characteristics on which its wrongness supervenes. It does not require the act of punishment to exhibit features of the offense that are irrelevant to its wrongness. Secondly, in specifying the wrong-making characteristics of the offense, we will be guided by our best sense of what made the action wrong in the first place. If all we have is an intuitive list of quite specific forbidden act-types, with no deeper explanation of their wrongness, then the list of punishments that LT requires will be similarly specific. If all our moral theory tells us is that it is wrong to knock teeth out and that it is wrong to gouge out eyes, if it offers no further account of *why* these assaults are wrong, then the only advice LT can offer is "An eye for an eye; a tooth for a tooth." But if we have a moral theory that is sophisticated enough to explain what is wrong about these superficially diverse instances of wounding, then that theory may provide the proponent of LT with a repertoire of possible punishments that is less crude than that required by the Mosaic injunction.

---

25. *See, e.g.*, SUSAN BROWNMILLER, AGAINST OUR WILL: MEN, WOMEN AND RAPE (1975). This suggestion is discussed and criticized by CATHERINE MACKINNON, FEMINISM UNMODIFIED: DISCOURSE ON LIFE AND LAW 6, 85-92, 233 n.19 (1987).

26. For a discussion, see Michael Davis, *Setting Penalties: What Does Rape Deserve?*, 3 LAW & PHIL. 61 (1984).

III

We face a choice in determining which of the wrong-making features of the offense should be reproduced in the punishment. A very strict version of *lex talionis* may require us to reproduce as many of those features as humanly possible, and it may require us to eschew any other principle or consideration in making this choice. Such a theory would need a very strong justification, for clearly there are other things that most people would want to take into account.

In the previous section, I mentioned that society is likely to be very uneasy with the infliction of any penalty which requires a sexual attack on an offender by an official.[27] The reason is not that the punishment is thought in principle inappropriate; it is our reluctance to encourage the nasty sadism and sexual corruption that such penalties would almost inevitably engender in the behavior of the officials involved. We just don't want our officials to be corrupted in this way. We do want to drive home to the sexual offender exactly what made his action (his rape or indecent assault) wrong, and subjecting him to some similar invasion, calculated to traumatize him and reduce his dignity, would be a way of doing that. But we have independent reasons for avoiding this, reasons that have to do with the moral integrity of the officials who have to administer the penal system. Thus we look for a more abstract description of the wrong-making features of the offense to reproduce in the punishment, or we resign ourselves to the reproduction of only some and not all of the wrong-making features in question.

A more relaxed version of LT might be able to accommodate this. If the principle is not seen as a categorical imperative, but as simply a rather good idea, then it would naturally take its place among a range of considerations (as other principles for determining amount and type of punishment do in more conventional penologies).

The corrupting effect of certain penalties is not confined to sexual cases. The prospect of sadism is likely to rear its head in any type of corporal punishment — flogging, for example. Such a prospect should give us pause in applying LT. While an assault for an assault seems a natural application of the principle, it may be that our reasons for not wanting officials to become assailants outweigh the reasons we have for preferring LT in this instance. If so, we may (regretfully) abandon the principle for such cases, or, again, move to some more abstract characterization of what is wrong about the offense of

---

27. Even Kant balked at the idea of raping the rapist:

> But how can this principle be applied to punishments that do not allow reciprocation because they are either impossible in themselves or would themselves be punishable crimes against humanity in general. Rape, pederasty, and bestiality are examples of the latter. For rape and pederasty, [the punishment is] castration (after the manner of either a white or a black eunuch in the sultan's seraglio), and for bestiality [expulsion forever from human society, because the individual has made himself unworthy of human relations.] *Per quod quis peccat, per idem punitur et idem.*

KANT, *supra* note 8, at 132. (I have supplemented Ladd's translation with that of W.G. HASTIE, KANT'S PHILOSOPHY OF LAW 224 (1887), because Ladd garbles the passage on bestiality.).

assault which can then be reproduced in the punishment without the potentially corrupting aspect.

Many similar considerations come together in the debate about capital punishment. Many who oppose execution as a penalty for murder do not do so on the ground that they think it in principle inappropriate, or that they think the murderer has an inalienable right not to be killed. Many of us oppose execution for reasons which have to do with its effects on the administration of justice and on the morale of the community as a whole, rather than its effects on the justly convicted criminal. Two considerations, in particular, are worth mentioning. First, in those few civilized countries which regularly practice capital punishment (for example, the United States and South Africa), there is an unhealthy preoccupation among members of the public with the grisly mechanics of execution, and a rather unpleasant display of gratification when someone is killed in this way. People are a little too eager that there should be executions, and a little too pleased when they happen. Some of us feel that these displays of official killing, pandering as they do to some rather nasty emotions in the community, are not entirely unrelated to the extraordinarily high homicide rate in the United States, for example. The unhealthiness of the preoccupation with execution, far from marking the society's revulsion from murder, seems to indicate a break-down in the taboos that are normally associated with the relishing of another person's death. The fact that, in the past, execution has had connotations of lynching and that there continues to be a racial aspect in the determination of who does and who does not suffer the death penalty, and the fact that judicial execution is still associated with the joys and vagaries, in the frontier mentality, of taking the law into one's own hands, are further aspects of this potentially brutalizing effect.

Less speculative, and perhaps more important, are considerations about possible miscarriages of justice. One of the things that had the greatest importance in the campaign for abolition of the death penalty in the United Kingdom was the occurrence during the 1950's of several famous executions of men who turned out later to have been innocent of the crimes for which they were put to death.[28] If a person turns out to have been innocent of some offense for which he was punished with a penalty other than death, there is something we as a society can do to make up for the suffering, loss of liberty, or inconvenience we have imposed on him. If he has been executed there is nothing we can do for him (though we may be able to compensate his estate or his dependents). Since we know that the penal system is imperfect, and that there are erroneous convictions from time to time, we may want a scheme of punishments which always leaves the opportunity to make some redress to those who turn out to have been wrongfully convicted. For this reason — because, as Charles Dickens put it, a penalty of death is "an irrevocable punishment [administered by] men of fallible judgment"[29] — we may want to strike execution from our list.

---

28. *See, e.g.*, Anthony Lewis, *Britain Pardons Man Hanged in 1950 for Murder of Daughter*, N.Y. TIMES, Oct. 19, 1966, at A19. For a full discussion of this case, see LUDOVIC KENNEDY, TEN RILLINGTON PLACE (1961).

29. Hugo Bedau, Miscarriages of Justice in Potentially Capital Cases, 40 STAN. L. REV. 21, 81 (1987) (citing CHARLES DICKENS, SELECTED LETTERS OF CHARLES DICKENS

Notice that nothing in this argument contradicts the suggestion, commonly associated with LT, that execution is an appropriate punishment for murder. Indeed, the very irrevocable nature of execution that is at issue here is precisely one of the things that makes it appropriate — for the victim of a homicide cannot be brought back to life or compensated either if the murderer repents or decides, on whatever grounds, that he has "made a mistake." The problem is that we want something in our penal practice to mark the fact that we are sometimes unsure that the person we are punishing is a murderer and that his action does warrant the infliction of death. Leaving open some possibility of redress to those who turn out to be innocent is one way of marking this. The moral credentials of such a reservation are sufficiently exhibited by its link with humility, moral fallibilism and an awareness of the limits of our own institutions; and they are thrown into relief by the arrogant confidence — and again, the corrupt enthusiasm for killing — that seems so characteristic of those who favor the death penalty.

So, to the extent that the proponent of LT faces a choice in his determination of which wrong-making features of the offense he should reproduce in the punishment, these considerations of corruption and uncertainty might play a role in guiding him away from reproducing the homicidal aspect of the offense.

It has to be remarked, however, that the arch-talon Kant would have none of this.

> If, however, he [the offender] has committed a murder, he must die. In this case, there is no substitute that will satisfy the requirements of legal justice. There is no sameness of kind between death and remaining alive even under the most miserable conditions, and consequently there is also no equality between the crime and the retribution unless the criminal is judicially condemned and put to death.[30]

What are we to make of this, in relation to what has just been said?

Note first that Kant goes straight on to insist: "But the death of the criminal must be kept entirely free of any maltreatment that would make an abomination of the humanity residing in the person suffering it."[31] Even if the criminal maltreated or tortured his victim, we must not maltreat or torture him, because that would corrupt our own moral character in regard to his humanity. This is an important recognition by Kant of exactly the *sort* of reasoning

---

215 (D. Paroissien ed., 1985)). That article contains an excellent survey of what is known about the incidence of capital miscarriages of justice. *See also* CHARLES L. BLACK, CAPITAL PUNISHMENT: THE INEVITABILITY OF CAPRICE AND MISTAKE (2d ed. 1981).

30. KANT, *supra* note 8, at 102. Hegel said something similar:

> Now although requital cannot simply be made specifically equal to the crime, the case is otherwise with murder, which is of necessity liable to the death penalty; the reason is that since life is the full compass of a man's existence, the punishment here cannot consist in a "value," for none is great enough, but can consist only in taking away a second life.

HEGEL, *supra* note 6, Addition to ¶ 101.

31. KANT, *supra* note 8, at 102.

we have been pursuing in this section.[32] That an offense involved torture is, in Kant's and most people's eyes, a morally relevant feature. Still, Kant insists quite rightly that we have good reasons — *self*-regarding reasons — for not wanting to reproduce this feature in the punishment.[33] He insists, however, that we may not make this choice about the infliction of death. So his position here depends on his view that killing someone (unlike torturing him) does *not* make an abomination of his humanity. Those who disagree with Kant about that will take the same attitude towards capital punishment as he takes towards judicial torture.

Kant evidently believes there is something *special* about killing. The fact that a death occurs, that there is a whole life lost, is a uniquely important feature of the murder, and crucial to our account of why it is wrong. He appears to think that we would be failing in our duty to visit on the murderer the universalized maxim of his offending if we did not put the murderer to death. Now someone might dissent from this view of Kant's in two ways.

First, he might deny that the fact that a person loses his life is the unique and all-important characteristic of a murder. He may concede its importance, but say also that murder has other features that make it wrong — for example, it contributes to fear and insecurity, it deprives dependents of their support and relatives of their loved one, it causes grief, it reduces the GNP, it disrupts whatever projects the victim was working on, etc.. All these seem to be wrong-making features which, though linked to the fact that death is involved, are, in principle, independent of that. Their independence is reflected in the fact that they or something like them could be reproduced in an act of punishment without killing the offender. For example, the imposition of imprisonment for life would share some of these characteristics with the offense of murder; and so it would respect LT to that extent.

Once again, those who think they know in advance of serious philosophy what LT requires will regard this as a mere gimmick, a wilful reinterpretation of the principle. But as we have seen, the principle necessarily requires that the act of punishment share *some features* with the act that constituted the offense. Which features these are will always be a matter of argument.

Secondly, Kant's opponent might concede that the killing aspect is all-important in the wrongness of murder, but he might propose — along lines discussed in the previous section — that we understand the wrongness of killing in terms that are a little more abstract in character. Why is killing

---

32. For further developments of this idea in the context of Kantian ethics, see Morris, *supra* note 15, at 270; Robert A. Pugsley, *A Retributivist Argument Against Capital Punishment*, 9 HOFSTRA L. REV. 1501 (1981). There is a critique of Pugsley's argument in Leon Pearl, *A Case Against the Kantian Retributivist Theory of Punishment: A Response to Professor Pugsley*, 11 HOFSTRA L. REV. 273 (1982).

33. DEATH IS DIFFERENT, *supra* note 1, at 41, argues that if we place constraints on the *type* of killing that we may inflict as punishment, then we are abandoning the strict principle of *lex talionis*: "Once we grant that such constraints are proper, it is unreasonable to insist that the principle of 'a life for a life' nevertheless by itself justifies the execution of murderers." Bedau is right that once we concede that *lex talionis* does not require us to reproduce the barbarity of the offense in the punishment, we cannot still say that it *requires* us to reproduce the homicidal character of the offense. But that does not mean that the principle may not *justify* execution, if we decide that killing *is* the feature we want to reproduce.

wrong? Because it radically disrupts an autonomous life. Very well, then let us radically disrupt the autonomous life of the offender. Does this mean we have to kill him? It depends on whether or not we have available some other punishment that shares this abstract feature with acts of killing.

Notice how in both limbs of this argument, the issue comes down to the way we characterize the wrongness of the offense. Do we say there are many things wrong with it or just one all-important wrong-making feature? And how do we describe such characteristics? The choice, as I argued earlier, is not a capricious preference for one description over another. It is a matter of *fit* with our best moral explanation of what makes the particular offense wrong. The principle itself, in its usual Mosaic formulations, does not resolve this issue for us: if it did, we should have to say that there is something irreducibly important about eyes and teeth! Rather, it is a principle which instructs us to base our choice of punishments on *our best understanding* of what is wrong with the offense.

IV

A common criticism of *lex talionis* is that it is incapable of explaining what counts as just punishment for offenses that do not involve the direct infliction of harm. The criticism goes as follows. Punishment is the direct infliction of harm, loss or suffering. LT requires that the act of punishment be like the act that constituted the offense. Therefore, LT applies only to offenses involving the direct infliction of harm, loss or suffering.

The flaw lies in the first premise. Defenders of LT regard themselves as offering a distinctive account of how people should be punished for the offenses they commit; they need not be bound by conventional views of what punishment must involve and, indeed, they are often critical of such views. Usually that criticism has to do with the type or amount of harm or suffering that is inflicted as punishment on the offender. But the LT critique may go beyond this. If the range of acts legitimately regarded as offenses extends beyond the class of acts which involve the infliction of harm or suffering, then the defender of LT will regard that as an indication that the range of punishments should be extended accordingly.

On the other hand, the proponent of LT, like anyone else, will have views on whether there should be victimless or harmless offenses. Those views may or may not be independent of his theory of punishment. It is sometimes thought that retributivists have difficulty in drawing a distinction between the set of wrong acts and the subset of punishable offenses, since retribution is supposed to be nothing but a response to moral wrongness. In fact, different retributivist theories offer different accounts of the point of punishment, and they therefore have some critical resources to sustain the distinction. Some aim to equalize benefits and burdens in the community, by offsetting the gains that people derive from offending.[34] Such theories will see no point in punishing immoral acts that do not upset the distribution of benefits and burdens. Other retributive theories, such as Hegel's, aim to annul viola-

34. *See, e.g.*, GEORGE SHER, DESERT 69-90 (1987); WOJCIECH SADURSKI, GIVING DESERT ITS DUE: SOCIAL JUSTICE AND LEGAL THEORY 225 (1985).

tions of right.[35] To the extent that I understand this theory, I believe the doctrine of annulment is supposed to turn crucially on the aspect of coercive injury involved in both crime and punishment. It follows that ethical lapses not involving coercion are not punishable according to the Hegelian approach.

In general, then: to say that a wrong act should be an offense under penal law is simply to say that punishment is appropriate for such a case. Whether one says that about a given wrong will depend on one's theory about the appropriateness of punishment. Since LT is usually the upshot of a theory of punishment, its proponent will have views on that as well.

However, in making that case, the proponent of LT need not be preoccupied with issues of direct harm and suffering. He will take LT as far as it can plausibly be taken, to determine who should be punished for what and how. If LT can only make sense of punishment for the direct infliction of harm or suffering, then that is where he will draw the line. But if the theory indicates a broader range of possibilities than that, he will draw the line somewhere else.

The feature which all defenses of LT have in common is that the criminal should experience an action which is, in some important respects like the offense he committed, from a point of view other than that of the agent. The infliction of suffering as punishment for the infliction of suffering is a particularly clear case of this. Having inflicted suffering on his victim and having therefore experienced such an action from the agent's point of view, the offender is now required to experience such an action from another point of view, namely, that of the victim. Our reason for requiring this may be educative or it may be seen, as it is in Kant's theory, simply as a requirement of reciprocity and universalizability. Either way, the assumption is that punishment puts the offender in someone else's shoes — that is, in some other role in an action of the type that constituted the offense.

But turning the tables on the offender, making him the direct victim of harm, is not the only way of changing roles. For example, Kant suggests the following as a punishment for theft:

> Inasmuch as someone steals, he makes the ownership of everyone else insecure, and hence he robs himself (in accordance with the Law of retribution) of the security of any possible ownership. He has nothing and can also acquire nothing, but he still wants to live, and this is not possible unless others provide him with nourishment. But, because the state will not support him gratis, he must let the state have his labor at any kind of work it may wish to use him for (convict labor), and so he becomes a slave, either for a certain period of time or indefinitely, as the case may be.[36]

Here the role that is thrust upon the offender is not that of the immediate victim of theft, but that of the indirect victim — the person whose ability to make a livelihood is imperilled by the insecurity of property rights. It may seem a far

35. *See* HEGEL, *supra* note 6, ¶¶ 91-103.
36. KANT, *supra*, note 8, at 102.

cry from "An eye for an eye, a tooth for a tooth." But we should recall that the act of theft has many features that make it wrong, that there are many roles in relation to those features, and that the application of LT necessarily involves some sort of choice as to which features we should reproduce, and which roles we should reverse, in the act of punishment.

In her book *State Punishment*, Nicola Lacey poses the following as one of two "devastating objections" which she says eliminate the principle ("if it merits the name") of *lex talionis* as an adequate explication of the retributive idea:

> Most obviously, in terms of the question of how much punishment should be inflicted, it supplies clear practical guidance as to the proper measure only in a selective number of cases. The penalty for murder or mutilation may seem clear, but what punishment ought to be inflicted for fraud, perjury or blackmail?[37]

It may be worth illustrating what I have said by sketching out an LT response for the offenses Lacey mentions.

An act of fraud has among its wrong-making features that it deprives a person of the information he needs to make a rational business decision, and that it contributes to general uncertainty and therefore enhances transaction costs in the business life of the community. We might respond to such an act along lines indicated in the passage from Kant just quoted: giving the offender a taste of the insecurity of livelihood that such challenges to the workings of an economy presage. Or we could give him a taste of economic uncertainty in other ways: for example, sentencing him to community service on days determined arbitrarily, unpredictably and at the last minute by a probation officer.

Perjury may or may not do harm to an assignable person, but it certainly impedes the ability of the courts to function. That is one of its wrong-making features. The offender has experienced that characteristic as agent, but what would it be like for him to experience it in some other role? Maybe a person guilty of perjury could be denied access to the courts for some specified period in the future. Or some property of his could be confiscated and returned to him (or not), depending on the toss of a coin. Such a punishment is unfamiliar, but LT makes no claim to model our "intuitions" of just punishment, if those "intuitions" have been based on some other principle or penal approach.

Blackmail presents the greatest problem of the three, partly because it is not at all clear why it is an offense in the first place.[38] (Again, remember that the proponent of LT is no more committed than the rest of us to the view that punishment is appropriate for an action like blackmail.) But suppose we agree that blackmail is wrong. Different explanations may be given; since these indicate different wrong-making features, they suggest, for the proponent of LT, different punishments. If the wrongness of blackmail has to do with the

---

37. NICOLA LACEY, STATE PUNISHMENT: POLITICAL PRINCIPLES AND COMMUNITY VALUES 17 (1988).

38. *See, e.g.*, Hugh Evans, *Why Blackmail Should be Banned*, 65 PHILOSOPHY 89 (1990).

invasion of privacy, the range of possible punishments is fairly obvious. A court officer might be appointed to go through the offender's personal effects and publish anything salacious he discovered. If the wrongness of the offense has to do with its exploitation of the victim's vulnerability, again it is not hard to imagine possible modes of retribution. The state has many ways of revealing a person's vulnerabilities and exploiting them.

Similar reasoning can be pursued for other cases. Herbert Morris offers this example of a just penalty for cheating in a children's game:

> [E]ven young children will find it particularly fitting to penalize a cheater by not permitting for a time at least, further play, for such a punishment conveys the central importance of honesty in the playing of the game and one's placing oneself outside the community of players by dishonesty.[39]

Here, as in Kant's theft example, the cheater changes roles not with the person most directly affected by his action, but with the generality of those who have something to lose directly or indirectly if his type of behavior becomes endemic.

This may seem a rather "creative" approach to punishment when compared to the stern rigidities that LT has traditionally been understood to dictate. But it involves nothing more original than (a) thinking through why the offense is wrong, and (b) thinking about the possibility of some of reversal of roles, which would allow the offender to experience an action relevantly like his offense from another point of view. These thought processes, (a) and (b), ought to be central to our moral reasoning anyway,[40] and it should be at least one of our aims (either in punishment or in general) to see that offenders undergo them.

V

In her book on punishment, Nicola Lacey also argues that *lex talionis* is incapable of explaining our sense that punishment should be varied depending on the degree of the offender's responsibility for the offense:

> [O]ur moral responses appear to differ enormously according to whether a killing is intentional or accidental; a wounding deliberate or negligent. This commonly acknowledged moral distinction between responsibility based merely on causation — strict liability — and that based on "mental elements" such as intent or recklessness — is ignored by the lex talionis, which directs the same response in each case.[41]

How can a principle which says such crude things as "A life for a life" accommodate itself to these distinctions?

---

39. Morris, *supra* note 15, at 266.

40. See especially the idea of role reversal in connection with universalizability in HARE, *supra* note 17, at 90-111. Hare has argued that moral judgments are, by virtue of their meaning, universalizable, so that one does not understand a judgment like "What Smith did to Jones was wrong" unless one understands what would be wrong about a relevantly similar act done by Jones to Smith.

41. LACEY, *supra* note 37, at 18.

The first thing to say is that LT is not put forward as a general principle for responding to death, blindness, dental injury, etc.. It does not hold that whenever someone dies, someone else must be killed, or whenever one tooth is lost, another tooth must be knocked out. It is put forward as a principle for determining *punishments* for *offenses*. This, by itself, should make us pause before accepting Lacey's imputation that the principle instructs us to treat accidental deaths on a par with intentional homicides.

Of course, as she notes, the principle does not provide us with a justification for punishment. But we saw in Section I that it is compatible with several justifications. Once these are brought into play, it is fairly clear how an approach that involves LT can deal with issues of responsibility.

Suppose the background justificatory theory is something like Kant's. It involves an attempt to bring home to the offender a universalized version of the maxim of his offense. Clearly it will be important for such an account, what the maxim of the original action was. A maxim such as "I will kill this person because his death will benefit me" differs considerably from "I will fail to take precautions against this person's death in performing a dangerous activity, because the precautions are inconvenient to me." No doubt a Kantian will judge them both wrong and punishable, but different arguments (different universalizing thought-experiments) will be offered in each case. If the point of LT is to cast the offender in some other role in the universalization of the maxim of his offending, then LT *will* recognize a distinction between intention and recklessness in regard to the causation of death and vary the punishment accordingly.

The same is true of a communicative theory of punishment, whether in a utilitarian or some other context. One of the things we want people to understand in society is that intentional malice is not the only evil; that recklessness and negligence are wrong also. At the same time, we want them to learn the moral differences among these various modes of wrongdoing. Though malicious and negligent injuring are both wrong, they are wrong in different ways. To drive home the wrongness of the former, we may deliberately and directly inflict suffering or injury upon the offender. But a proponent of LT may well regard that as inappropriate for the punishment of negligence. Instead, he may propose subjecting the offender to various forms of fear or insecurity (along lines sketched out in Section IV), or he may propose some sort of lottery, with the infliction of injury as a possible, but not inevitable, consequence.[42]

What about insanity? Here it seems LT can accommodate most of our intuitions. If a criminal was so insane at the time of the offense as to be not responsible for his action, then there will be no point, either on a Kantian theory or on a communicative theory, for holding him liable to punishment. In such a case, there is no *action of his* to re-enact in the way of punishment; there is only an arbitrary piece of behavior generated pathologically, and the proponent of LT will see no point in recreating that for the purposes of moral

42. *See* David Lewis, *The Punishment that Leaves Something to Chance*, 18 PHIL. & PUB. AFF. 53 (1989). This approach could, as Lewis urges, be used in punishing attempts, and also for punishing actions like drunk driving which are condemned on account of their dangerousness rather than on account of any injury actually caused.

education or for the sake of universalizability.[43] In addition, communicative theories can make perfect sense of our reluctance to punish those who have become insane since their offending.[44]

The only aspect of our modern practice that might embarrass a proponent of LT is the quantitative diminution of responsibility — the idea that certain mental states (provocation, for example, or minor disturbances or inadequacies) diminish responsibility without exculpating the offender completely. Other retributivists pursuing theories of desert may have less difficulty with this. If the aim of punishment is to confer an amount of suffering equal or proportional to the degree of wrongness of the offense, then we may use the formula suggested by Robert Nozick — "r x H," where "H" represents the magnitude of the wrongness of the act done by someone fully responsible, and "r" ($0 \geq r \geq 1$) represents a particular offender's degree of responsibility for the action.[45] But LT is incompatible with such a quantitative idea of desert. LT dictates a qualitative approach to punishment, one in which the leading idea is, as we have seen, the *similarity* of punishment to offense, not its equality or proportionality.[46]

The quantitative conception of diminished responsibility is, in any case, problematic and controversial, and it may not count against LT that it cannot explain it. LT can certainly explain why offending under provocation might differ, in its appropriate punishment, from premeditated offending. And it may be able to explain the significance attached to minor mental disorders. But it cannot offer a quantitative framework to accommodate these points.

VI

In concluding, I want to ask whether the various steps in this argument have changed or weakened the principle of *lex talionis* out of all recognition. In an effort to make it plausible, have I not in effect turned it into something else?

The Essay has indicated three moves that might be thought to weaken or qualify LT. First, in Section II, I insisted that the principle requires the choice of a punishment which is similar rather than identical to the act that constituted the crime. I argued that this provided an opportunity for selecting

---

43. *See* IMMANUEL KANT, *De Imputatione*, *in* LECTURES ON ETHICS 57 (Louis Infield trans., 1963):

> To impute responsibility is to judge, in accordance with certain practical laws, how far an action is due to the free agency of a person.... We can ... attribute certain actions to a lunatic or to a drunken man, but we cannot hold them accountable for them, for they are not free agents. The drunkard can, to be sure, be held responsible for his drunkenness when he is sober, but not for his actions when in drink.

44. *See* ROBIN A. DUFF, TRIALS AND PUNISHMENTS 14-38 (1986).

45. NOZICK, *supra* note 14, at 363.

46. Proportionality may or may not be a byproduct of the application of *lex talionis*. In principle, a defender of *lex talionis* has no interest in establishing a scale of offenses and penalties ranked according to gravity and severity, respectively. His approach to punishment does not require or presuppose it. I think this is an advantage, since the linear rankings of gravity and severity required by other retributivist theories always seem artificial. For the best discussion of such rankings, see Michael Davis, *How to Make the Punishment Fit the Crime*, *in* NOMOS XXVII: CRIMINAL JUSTICE 119 (J. Roland Pennock et al. eds., 1985).

which features, among all the various features of the offense, we should attempt to reproduce in the punishment. Second, in the same section, I argued that we have to make a choice about the level of abstraction at which similarity is to be established. Third, in Sections I, III and V, I suggested that LT might be only one consideration among several to be taken into account in our choice of punishment.

Let me say something about the third of these moves. In effect, it amounts to a claim that even if LT is an important principle, it may have less than absolute weight in guiding our choices. Our choice of punishment will also be guided by other, and at times competing, principles like social cost, fallibility, the need to avoid the corruption of penal officials, and the other aims and goals of the penal system. Now this sort of compromise certainly represents a move away from LT understood as an absolute principle of retribution. But no one has ever been able to provide a defense of it as an absolute principle. The two best-known justifications are, as we saw, the Kantian approach and the communicative approach to punishment. But Kant's system requires at least some compromise in order to avoid "abominations" that might undermine the much deeper principles of respect for humanity that his approach to punishment presupposes. And the communicative model, aiming as it does at moral education, will at the very least require compromises in relation to other principles of education that might from time to time compete with the application of LT. Though LT may be a rather good idea for bringing home to offenders the true nature of their offending, it is unlikely to be the only good idea that we have in the area. By its very nature, a theory of moral education is always liable to be compromised against some of the principles and goals — of compassion and consideration, for example — that we are using it to inculcate.

I find it increasingly difficult to think what people could have in mind when they insist that LT has to be understood as a moral absolute. To justify any principle is usually to appeal to deeper, more foundational considerations. They need not be consequentialist considerations, but there must be deeper justifications nonetheless. It is always possible that those deeper considerations will provide a basis for other principles as well as the one we are trying to justify. So respecting the deeper considerations may unavoidably involve an attempt to balance its different implications against one another. The only way to avoid this would be to insist that the principle in question — in this case, LT — is itself a first principle of moral justification, to insist, in other words, on its axiomatic character or its intuitive self-evidence, and to deny that it is supported by any deeper principle capable also of supporting other ideas that might compete with it. Although some may have intended the principle in this light, it is hardly a candidate for our most charitable understanding of *lex talionis*.

Even if a principle is less than absolute, it may still *make a difference* to the moral calculations in which it is considered. An approach to punishment in which LT is balanced alongside other penal aims will differ somewhat from an approach in which LT plays no part at all. In order to see what part it plays, even in such a moral balancing, we have to understand what it means. My enterprise of trying to work out what LT entails, of trying to see what it means to reproduce the crime in the punishment, is perfectly compatible with —

indeed it is presupposed by — my insistence that the principle has only *prima facie* status. We couldn't even think what moral weight to accord to it unless we knew what it implied.

Moreover, I have not simply asserted that LT is a *prima facie* principle, and left it at that. I have indicated *ways* in which it might be balanced with other principles and concerns.[47] Since LT leaves some room for choice of punishments by indicating that we have to make decisions about which features of the crime we want to reproduce, we are provided with a technique for involving other morally relevant considerations in our deliberations on this matter. If we are worried, for example, about the corruption or demoralization of penal officials, we can direct legislators to choose features which are least likely to have this effect while still serving the point of LT.

The other two points directly concern the choice of features. I have argued that choice is unavoidable since the crime does not itself *tell us* which features of it should be reproduced in the punishment. We want the punishment to be similar to the crime, but it is up to us to indicate both the respects in which it should be similar and the level of abstraction involved in determining similarity. To say this is not to say that we can choose anything we like. Once again we must look to the justification of the principle, rather than thinking we can simply "read off" what it requires from the ten characters that make up the phrase "*lex talionis*" or from a few scattered verses in the Pentateuch. Both the Kantian approach and the moral education approach indicate constraints on our choice. Both indicate that we should choose the important and relevant features of the offense on which the best understanding of its wrongness supervenes.

That this does not amount to a precise algorithm stems from the fact that the wrongness of many offenses is a complex matter and people disagree about how to understand the various aspects of that complexity. Feminists disagree about how exactly to understand the wrongness of rape. Kantian and utilitarian moralists disagree about how to conceive the wrongness of murder. And so on. These disagreements are simply there (though that is not to deny that there may be better or worse answers) and their existence cannot be blamed on the proponent of LT. Anyway, on the interpretation I have given, the implications of LT cannot possibly be more determinate or less controversial than our understanding of the wrongness of the offenses we are punishing.

Although my aim in this Essay has been to reinterpret LT and to provide a better understanding of the various dimensions of its complexity, I have tried to avoid turning it into any other principle of punishment. The obvious temptation is to reinterpret LT as a quantitative principle of desert, along lines suggested by Hegel. In his discussion in *Philosophy of Right*, Hegel repudiated any theory of superficial equality between crime and punishment: "theft for theft, robbery for robbery, an eye for an eye, a tooth for a tooth — and then you can go on to suppose that the criminal has only one eye or no teeth." Instead, he said we should look for a "deeper" equality of "value":

---

47. Following the approach urged in Jeremy Waldron, *Rights in Conflict*, 99 ETHICS 503, 516 (1989).

> In crime, as that which is characterized at bottom by the infinite aspect of the deed, the purely external specific character vanishes all the more obviously, and equality remains the fundamental regulator of the essential thing, to wit the deserts of the criminal, though not for the specific external form which the payment of those deserts may take. It is only in respect of that form that there is a plain inequality between theft and robbery on the one hand, and fines, imprisonment, &c., on the other. In respect of their "value", however, i.e. in respect of their universal property of being injuries, they are comparable.[48]

To Hegel it seemed that the only alternative to a superficial equality involving eyes and teeth was some attempt to equalize a homogenized degree of "injury" involved in the offense with a similarly homogenized degree of "injury" involved in punishment. Now this is an interesting theory, but it has little in common with, and could not be sustained by, the justifications of LT that we have discussed. For, on both the Kantian and the moral education model, the punishment was supposed to provide an indication of how (i.e. in what way) the offense was wrong. On the Hegelian approach, by contrast, all punishment provides is an indication *that* the offense was wrong, and a quantitative indication, either directly or proportionately, of how wrong it was.

The other alternative I wanted to avoid was the approach that identifies retribution with the equalization of benefits and burdens in society. According to this view, the criminal deserves punishment because in offending he has cast off a burden of self-restraint that all of us are required to bear and so disrupted the fair distribution of the costs of social life. An appropriate punishment is one that offsets this stolen advantage, and restores the original distribution.[49] Murphy has maintained that this was, in fact, Kant's view, but I am afraid he is mistaken: Kant says nothing to warrant that interpretation.[50] His theory is based, as we have seen, on ideas connected with universalization rather than on ideas associated with distributive justice.

In any case, LT is quite different from the benefits-and-burdens view. As we have seen, it prescribes that the criminal should suffer as punishment the performance of an action which is like his crime in certain respects, but performed with him cast in some role other than that of agent. Suppose, to take the clearest case, we cast the criminal as the victim of the action: previously he was the assailant, now he is the person assaulted. In this case, he will suffer something similar to what his victim suffered. Now that similarity is quite a different matter from any attempt to offset the advantage he derived from the assault. For all we know, the advantage he gained (the indulgence of giving in to anger, for example) is much less than the suffering of the person he struck. There is no reason to think that what is necessary to offset the criminal's advantage is in any way identical with, or equal or similar to, the amount of suffering that a victim of the offense would experience. Since LT requires us to try and duplicate the latter experience for the offender, it cannot

48. HEGEL, *supra* note 6, ¶ 101 (Remark). *See also* IGOR PRIMORATZ, JUSTIFYING LEGAL PUNISHMENT 80-81 (1989).

49. *See* SHER, *supra* note 34, at 91; SADURSKI, *supra* note 34, at 225.

50. JEFFRIE G. MURPHY, KANT: THE PHILOSOPHY OF RIGHT 142-43 (1970).

be interpreted as a theory that aims specifically to offset the offender's unfairly gained advantage.[51]

51. George Sher, however, insists that the unfair advantage gained by the criminal is always, in fact, proportional to the suffering or harm done to his victim. He suggests that the criminal benefits by escaping from a constraint, and that the more important the constraint the greater the benefit. Since the importance of the constraint corresponds, on Sher's account, to the harm to a potential victim, the benefit to the offender will accordingly also correspond to the victim's harm. *See* SHER, *supra* note 34, at 74-90. But there's a fallacy in this argument. Though the *seriousness* of the offender's violation is determined by the harm that is done, that consideration does not determine how much he stands to *gain* by harming his victim. One may gain a lot by inconveniencing someone slightly or gain a little by the infliction of the gravest harm. See also JULES L. COLEMAN, MARKETS, MORALS AND THE LAW 187 (1988), for similar points about tort law.