Punishment Theory's Golden Half Century: A Survey of Developments from (about) 1957 to 2007

Article *in* The Journal of Ethics · March 2008

DOI: 10.1007/s10892-008-9040-0

1 author:

Michael Davis
Illinois Institute of Technology
206 PUBLICATIONS **2,236** CITATIONS

SEE PROFILE

J Ethics (2009) 13:73–100
DOI 10.1007/s10892-008-9040-0

Punishment Theory's Golden Half Century: A Survey of Developments from (about) 1957 to 2007

Michael Davis

Received: 21 April 2008 / Accepted: 1 May 2008 / Published online: 10 September 2008

Abstract This paper describes developments in punishment theory since the middle of the twentieth century. After the mid–1960s, what Stanley I. Benn called "preventive theories of punishment"—whether strictly utilitarian or more loosely consequentialist like his—entered a long and steep decline, beginning with the virtual disappearance of reform theory in the 1970s. Crowding out preventive theories were various alternatives generally (but, as I shall argue, misleadingly) categorized as "retributive". These alternatives include both old theories (such as the education theory) resurrected after many decades in philosophy's graveyard and some new ones (such as the fairness theory). Only in the last decade or so have new vares o "consequentialism" appeared to dilute a debate among philosophers that had become almost entirely about "retributivism". I shall describe this trend in more detail. The description will be less an update of my 1990 survey than a rethinking of it. The conclusion I draw from this rethinking is that we need to drop the utilitarian–retributivist (and nonconsequentialist–nonconsequentialist) distinction in favor of one sorting punishment theories according to whether they rely in part on empirical considerations (externalist theories) or instead rely (almost) entirely on conceptual relations (internalist theories).

Keywords Retributive · Utilitarian · Consequentialist · Nonconsequentialist · Conceptual · Empirical · Kant

1 Introduction

Punishment theory has both prospered and changed in the last half century as it had not in a long time—if ever before. We are in, though perhaps at the end of, a golden

M. Davis (✉)
Department of Philosophy, Illinois Institute of Technology, Chicago, IL 60616, USA
e-mail: davism@iit.edu

age. No doubt this is part of larger events, the rise of practical philosophy in general and legal philosophy in particular, but no less real for that.[1]

The overall trend in theory is also clear. After the mid-1960s, what Stanley I. Benn called "preventive theories of punishment"—whether strictly utilitarian or more loosely consequentialist like his—entered a long and steep decline, beginning with the virtual disappearance of reform theory in the 1970s.[2] Crowding out preventive theories were various alternatives generally (but, as I shall argue, misleadingly) categorized as "retributive".[3] These alternatives include both old theories (such as the education theory) resurrected after many decades in philosophy's graveyard and some new ones (such as the fairness theory). Only in the last decade or so have new varieties of "consequentialism" appeared to dilute a debate among philosophers that had become almost entirely about "retributivism". I shall describe this trend in more detail. The description will be less an update of my 1990 survey than a rethinking of it.[4] The conclusion I draw from this rethinking is that we need to drop the utilitarian–retributivist (and consequentialist–nonconsequentialist) distinction in favor of one sorting punishment theories according to whether they rely in part on empirical considerations (externalist theories) or instead rely (almost) entirely on conceptual relations (internalist theories).

Some surveys of punishment theory over the last half century include several topics I shall ignore here. Two, determinism and free will, are generally no longer considered part of punishment theory. The other topics I shall ignore still are (more or less) part of punishment theory: what offenses it is appropriate to punish ("the limits of the criminal law"), states of mind that reduce criminal responsibility, excuses generally, clemency, and prisoner's rights. I shall ignore them not because they do not deserve careful discussion but because I cannot discuss them carefully here while keeping this article to a polite length.

[1] Consider just the number of entries under "punishment" in the *Philosopher's Index* for each ten year period, beginning with 1957 (the first decade of the half century): 104 for 1957–1967; 503 for 1997–2007 (a five-fold increase). The numbers of items for the intervening decades are: 224 for 1967–1977 (a doubling); 297 for 1977–1987 (small increase); 370 for 1987–1997 (increased by a third—the same as for the most recent decade). All of these numbers are, of course, impressive compared to the *Index*'s first decade, 1940–1950, which had only 30 items. For that reason, as well as because of the inventiveness of the decade beginning with 1954, a poetic license seems unnecessary to claim a golden *half century* (give or take a few years). Of course, the *Philosopher's Index* understates the actual literature in question, since a significant part appears in legal journals and other academic publications the *Index* does not index. These numbers are, only suggestive; alone, they prove nothing about the quality of the literature in question.

[2] Edwards (1967, pp. 29–36).

[3] See, for example, Dolinko (1991), which begins with several paragraphs describing the rise of retributivism both in practice and in theory. Dolinko is among the few remaining utilitarians doing punishment theory—and one of the best critics of retributivism.

[4] Davis (1990).

2 Defining the Subject: 1954–1967

The decade or so before Benn's classic entry on punishment in the *Encyclopedia of Philosophy* was, in general, a bad time for fields of philosophy concerned with "values".[5] Analytic philosophy—which in its twin forms, logical positivism and linguistic analysis, dominated philosophy in the English-speaking world—seemed to regard "value talk" as something between an intellectual mistake and a temper tantrum. It was nonetheless during this period, and thanks to analytically minded philosophers, that punishment theory was given two tools that seem to have made possible subsequent developments.

One of the tools was a definition of punishment that allowed philosophers to distinguish clearly between criminal punishment, always the actual (or, at least, primary) subject of punishment theory, and various other topics which, though using the term "punishment" (in a recognizably related sense), seem to belong to other fields: divine punishment, the punishment of young children, the punishment of animals, natural punishment ("poetic justice"), revenge, and so on. We might summarize the various statements of the definition in this way:

Punishment theory is (primarily) concerned with an institution having:

1. A body of rules (and principles for interpreting them) capable of guiding ordinary conduct, what we may call "primary rules";
2. Rational agents, that is, beings capable of following these rules or not as they choose, capable of choosing on the basis of reasons, and capable of treating the prospect of an undesirable consequence, even one distant in time, as a reason against doing an act (to be weighed with other reasons for and against);
3. Secondary rules designed to connect (and generally succeeding in connecting) failure to follow primary rules with penalties (that is, with specified undesirable consequences);
4. Conventional procedures for imposing penalties upon rational agents in accordance with the secondary rules;
5. A justified presumption that both primary and secondary rules (especially rules setting penalties) are generally known to those subject to them; and
6. A practice of justifying imposition of a penalty (in part at least) by the fact that the individual upon whom it is to be imposed, though (more or less) rational, failed to follow the appropriate primary rule.

This is often called the "Flew-Benn-Hart definition" because Benn, Anthony Flew, and H.L.A. Hart all defended (something like) it more or less independently at about the same time.[6] I shall hereafter call it "the standard definition" (both for brevity and because that is what it has become).

[5] Edwards (1967, pp. 29–36).

[6] See Flew (1954), Benn (1958), Hart (1960). For a sustained critique of this definition, see McCloskey (1962)—which denies, among other things, that the standard definition identifies a central "problem of punishment" the solution to which is likely to be independent of the solution of any corresponding problem where "punishment" has one of its other senses; and also McPherson (1967), which denies, among other things, that "punishment" can usefully be defined except within a specific theory of punishment.

Because the standard definition requires punishment to be by a conventional procedure (4) and to use specific penalties known in advance (2, 3, and 6), divine punishment of the traditional Christian sort must be treated as only an analogue of punishment proper. What we know of divine punishment, if we know anything, is that it will be so bad that, having it clearly before us, we would choose to act as God wishes. There is an ancient, but still lively, debate about what divine "punishment" is (regret, separation from God, physical torment, or something else).[7] Perhaps we can say everything we might want to say about "the wrath of God" without using "punishment" (or "penalty").

For the standard definition, punishing a child or animal also can only be an analogue of punishment proper for much the same reason divine punishment is. Generally, children, especially young children, and animals are not told in advance what the penalty for breaking a rule is, and there is often (especially at home) no conventional procedure for imposing the penalty, just an individual act (a parent deciding "on the spot"). Within wide limits, both parents and the owners of animals may "discipline" their charges in any way they decide. In this respect, parents and owners are little gods.

There is, however, also a reason to think the discipline of children or animals is even less like punishment proper than divine punishment. Insofar as we recognize a being as a mere child or mere animal, we recognize it as less than rational—that is, as at least largely lacking the ability to follow rules, to choose on the basis of reasons, and to treat the prospect of a specific penalty as a reason to follow the rule. However discipline of children or animals is to be justified, it cannot be justified *as punishment* (or, at least, as punishment proper).

The other analogues of punishment—natural punishment, revenge, and so on—fall short of counting as punishment proper for these (and perhaps other) reasons.[8]

Once punishment theory was defined as concerned with an institution satisfying conditions 1–6 (that is, as primarily concerned with criminal law), it was obvious (or, least should have been) that the purpose of punishment theory, "the justification of punishment" or (as analytic philosophers preferred) the analysis of its justification, could answer one or more of several different questions. Initially, only two questions were distinguished: a) the ("moral") question of *justifying a rule, practice, or institution* of punishment; and b) the ("logical") question of *justifying an act* under the rule, practice, or institution.[9] In time, though, philosophers identified at least six distinct questions a (complete) theory of punishment could (and, indeed, should) answer (or, at least, provide a framework for answering):

[7] See, for example, Cain (2002), Walls (2004), Kabay (2005).

[8] Note, for example, how a recent attempt to link punishment and revenge begins with a definition of punishment ignoring many of the features of the standard definition. Zaibert (2006). The assumption is that punishment is a practice possible in the state of nature. Yet, the origin of the word "punishment" is legal (or at least institutional). To assume the possibility of punishment in the state of nature seems to be like assuming a military order or tax bill in the state of nature; the assumption must rest on confusion or mere analogy. Institutionality is one of the ways in which punishment differs substantially from mere harming in response to wrongdoing.

[9] The modern discussion of this distinction seems to have begun with A. M. Quinton's article, "On Punishment", *Analysis* 14 (June 1954): 133–143.

1. *The philosophical question*. Why (and under what conditions) establish any institution of punishment at all (an institution being an organized body of rules, offices, and practices)?
2. *The political question.* Why (and under what conditions) establish this institution with its special concepts, principles of legislation, adjudicative procedures, and permissible penalties rather than some other?
3. *The legislative question.* Why (and under what conditions) assign this penalty for this sort of rule violation rather than some other penalty (or none at all)?
4. *The eligibility question.* Why (and under what conditions) assign this punishment to this criminal (that is, impose this penalty) for this rule violation (or, if the criminal is to be let go, why do that)?
5. *The sentencing question.* Why (and under what conditions) assign this punishment to this criminal rather than more or less (or some other sort)?
6. *The administrative question.* Why (and under what conditions) carry out the assigned punishment rather than some other (or none at all)?

While the greatest part of work on punishment is and, indeed, always has been concerned primarily with the first question, philosophers have often dealt with more than one and, in the last few decades, have increasingly concerned themselves with the last four. There are now substantial literatures concerned with proportion in statutory penalties and individual punishment (3 and 5), responsibility and excuse (4), and clemency (both judicial, 5, and executive, 6).

There are at least two reasons for this interest. First, insofar as theories help us to see phenomena more clearly, a theory of punishment should help us resolve practical problems of punishment more easily. Application of a philosophical theory to the increasingly practical questions 2–6 often reveals important weaknesses (and, occasionally, strengths) of the theory. Second, the practical questions often involve philosophically interesting issues separate from punishment's philosophical question (1). So, for example, clarifying the concept of legal responsibility has often suggested ways to clarify the related concepts of moral responsibility, personal responsibility, and collective responsibility.

The analytic philosophers who first drew the two-part distinction did so to resolve the contest between utilitarian and retributivist theories of punishment by admitting each theory to be partly right (and partly wrong), but they always gave utilitarian theory the more important role. So, for example, John Rawls assigned utilitarian considerations the role of justifying the practice of punishment (answering question 1), leaving retributivism to justify penalties under the practice (to say, in effect, "follow the rules of the practice" in answer to questions 4–6).[10] Though popular for a few years, this way of resolving the dispute (the classic "mixed view") soon led to restatements of retributivism that, while acknowledging the distinction between questions, offered reasons for thinking that utilitarianism could not provide the justification required even for the rule, practice, or institution while retributivism

[10] Rawls (1955, p. 5): "utilitarian arguments are appropriate with respect to questions about practices while retributive arguments [wrongdoing merits punishment and punishment should be proportioned to wrongdoing] fit the application of particular rules to particular cases." Of course, Rawls himself gives a long list of precursors for his mixed view. Rawls (1955, p. 3, n2).

could. I shall discuss some of these reasons in the next section because they largely explain the great decline of utilitarian theories of punishment. But I must explain one reason now because it concerns the definition of the problem that a theory of punishment should solve.

For each of the six questions we just identified, a theory of punishment is supposed to provide a *justification* (or, at least, to analyze the conditions under which a justification can be given). A justification is a set of propositions designed to show that a certain course of conduct (or belief) is rational. We may distinguish at least four senses of "justification" in which a theory might justify punishment (that is, answer the philosophical, political, legislative, eligibility, sentencing, or administrative question). A justification might show (a) that punishment is *morally permitted* (rational, all else equal), (b) that it is *morally required*, (c) that, though only morally permitted, it is, all things considered, still something *reason requires (*that is, merely rationally required*)*, or (d) that, though morally permitted but not morally or merely rationally required, still is something that reason recommends (that is, that the conduct in question is not only all right but positively good).[11] Theorists offer a justification of the first sort when, asked to justify punishment, they appeal to "our right" to punish. The right makes punishment morally permissible (without making it good or required); it does not so much answer the question "Why should we do it?" as "Why is it okay to punish?" Theorists offer a justification of the second sort when, say, they argue that punishment is a duty of justice. Theorists offer a justification of the third sort if they argue that we should punish because we have the right to punish and a pressing practical need to exercise that right (for example, because punishment is the only way to keep crime in check). Circumstances leave us "no choice" about how to use our right. Theorists offer a justification of the fourth sort if they argue that punishment is not only morally permissible but good because it serves some moral ideal (such as desert) or some merely rational ideal (such as efficiency). We may, of course, be free to serve another ideal instead even if that ideal does not involve punishment. Since each of questions 1–6 can have at least four answers (four different justifications a–d), a complete theory of punishment consists of up to 24 (more or less related) elements.

Utilitarianism has a tendency to show that conduct is required (or forbidden)—justified in the second or third sense. This is as true of utilitarian theories of punishment as of utilitarian moral theories. Mere permissibility or positive goodness that does not require is a problem within utilitarianism. Under an act-utilitarian theory, nothing is merely permitted except when its utility and that of the best alternative are equal. But even rule-utilitarians can only recognize moral permissibility if their theory somehow limits the number of moral rules (and what they require). If moral rules cover too much, there is again no mere morally permissible or merely morally good conduct. Utilitarian theories of punishment (though

[11] The distinction between the morally permissible and the rationally required seems to be another important development of the decade before Benn's *Encyclopedia* entry (Edwards 1967, pp. 29–30). See Armstrong (1961, p. 474), which distinguishes between the "point" of punishment (what makes it rational) and its "justification" (what makes it morally permissible). This article opens with an extended description of the marginal status of retributive theory at that time, a striking reminder of how much the second half of twentieth-century punishment theory differs from the first.

logically distinct from utilitarian moral theories) have much the same tendency to demand too much. They therefore almost never try to offer any justification but b.

Any theorist adopting (something like) a classic mixed view therefore confronts a dilemma: (1) embrace utilitarian moral theory and risk having the mixed view collapse into a purely utilitarian theory of punishment; or (2) embrace a non-utilitarian moral theory (for example, one recognizing the claims of justice as independent of utility) and risk having non-utilitarian considerations (such as justice) pre-empt utility even when justifying punishment as an institution.

Retributivism, in contrast, may (and often does) presuppose a morality consisting of a small number of relatively undemanding constraints, with perhaps some ideals (morally worthy targets). For such a theory, morality can be much more about defining what conduct is permitted or good than about what is required. A retributive theory might, then, understand the justification of punishment in the morally permitted sense. We have many good reasons to punish. But such reasons—the functions, aims, rationales, or points of punishment—cannot justify conduct that morality forbids. That seems as true at the level of institutions as at the level of individual acts. Since there is no reason why many of the "side-constraints" that apply to the choice of individual acts could not apply as well to the choice of rules, practices, or institutions, a retributivism of side-constraint could threaten the classic "mixed view".

There is, however, another sort of mixed view that a retributivism of side-constraints does not threaten. This is a mixed view in which retributivism provides the side-constraints and utilitarian theory (or, at least, consideration of contingent consequences) provides the reasons for acts within those side-constraints. Given that we have the right to punish (that is, the right to establish an institution of punishment as well the right to punish this individual within that institution), the mixed theory would identify the contingent consequences (say, deterrence) that justify (make it good) to punish. Several writers have attributed such a mixed view to Kant.[12] I regard this attribution as a mistake, though one involving an important insight into Kant's theory. I will explain why later.

Those who find a side-constraint version of punishment theory necessarily too weak to be interesting should consider the context in which the problem of punishment arises. Punishment is not a merely *possible* institution someone has suggested or even an institution (like polyandry) limited to a few societies. Punishment has an important place in every substantial legal system. Few people doubt that punishment is sometimes useful or even necessary. What seems to worry most of those whom punishment worries at all is how punishment coheres with morality. Moral rules generally forbid killing, maiming, holding others against their will, taking the property of others without their consent, and causing pain; yet, punishment (apart from suspended sentences) consists almost entirely of such acts. A theory of punishment that merely showed that punishment (perhaps somewhat revised) is morally permissible in (something like) existing conditions would resolve that worry. While some people, perhaps most, would also like guidance concerning how to exercise the right to punish once justified, that guidance seems to answer (partially or completely) questions 2–6 rather than the first question. The

[12] See, for example, Merle (2000), Brooks (2003), Tunick (1996), Scheid (1983).

deep question, the philosophical, has been answered just by showing that punishment is morally permissible. A more demanding answer is unnecessary. Theorists need to be clearer than they generally are about the question they are answering (1–6 or some other), the sort of justification they are giving (a–d or some other), and why they are answering that question.

3 Prolegomenon to Utilitarian (and Other "Consequentialist") Theories

Utilitarian theories of punishment are, it is said, "consequentialist", that is, they attempt to justify punishment *solely* by its "consequences".[13] I have italized "solely" because most theories of punishment, not just the classic mixed views, take account of consequences in some way. Even Kant, that strictest of retributivists, seems to admit that, *all else equal*, it would be permissible to choose a penalty that serves some utilitarian purpose (for example, saving money) over one that does not.[14] What distinguishes consequentialist theories properly so called, especially utilitarian theories, from the others, the so-called "nonconsequentialist theories", is that consequences (directly or indirectly) are *all* that matters. Consequentialists are, by definition, "one-note Johnnies".[15]

Or, rather, that would be what distinguishes them did the term "consequences" not have at least two senses that punishment theory needs to keep distinct. In one sense, the logical, we can say that the conclusion of an argument is the consequence of the antecedent premises. The connection between antecedent and consequence is "internal" to the concepts involved (and eternal in the way logical relations are).[16] So, for example, *4* is the (logical) consequence of doubling *2* (always and

[13] Utilitarian theories of punishment seem differ from other consequentialist theories either in what consequences count or in how they are counted. For example, a utilitarian cannot (without ceasing to be utilitarian) give up the idea of maximizing good consequences (however defined) or count justice or equality as a consequence (that is, as good in addition to human happiness, social welfare, or whatever is the measure of utility). Nothing in what follows depends on this way of distinguishing utilitarianism from other forms of consequentialism—except the form of exposition.

[14] My case for that claim consists, in part, of a single word "merely" in the following sentence: "Judicial punishment can never be used merely as a means to promote some good for the criminal himself or for civil society". Kant (1999, p.138). The other part of my case for the claim, too long to give here, is the congruence between the interpretation I have given this passage and various standard interpretations of Kant's second version of the Categorical Imperative. Others have reached the same conclusions for somewhat different reasons. See, for example, Corlett (2006, p. 56). While Corlett seems to regard this as a significant departure from "pure retributivism", I do not. It is a tie-breaking procedure. Such procedures typically depart from whatever procedure led to a tie in only a few cases and only after the procedure has done all it could. The justification of a tie-breaking procedure is (a) that some decision is better than no decision and (b) the procedure is convenient without sacrificing anything of significance.

[15] Many, perhaps most consequentialists, also "maximize" the consequence in question (differing in the good they seek to maximize and the costs they deduct). But nothing prevents a consequentialist from aiming at something less or different. And, indeed, any theory that recognizes a plurality of incommensurable goods (or evils) must adopt a standard of evaluation falling short of simply maximizing the good.

[16] This use of "internal" goes back at least to Hegel (2008), for example, Introduction §2:

> The science of right is a part of philosophy. Hence it must develop the idea, which is the reason of an object, out of the conception. It is the same thing to say that it must regard the peculiar internal development of the thing itself.

everywhere)—though, empirically (for example, because of error), the consequence (of the operation of doubling) might not be *4*. Most punishment theories (and perhaps all) are *in part* consequentialist in this strictly logical sense (that is, insofar as each relies on deductive argument). Many theories of punishment are consequentialist in a somewhat looser (but still conceptual) sense. They include claims about what we do by doing something else (where one act includes the other), for example, that one consequence of punishing the guilty is that justice is done (that is, by the very act of punishing the guilty we simultaneously do justice). Consequentialist theories properly so called are concerned with consequences in a quite different sense, that is, with *events* consequent on other events merely as a matter of fact (where "fact" includes contingent scientific laws). The argument for a consequentialist theory must depend in part on how the world happens to be.

For this reason, sorting theories of punishment into "consequentialist" and "nonconsequentialist" is misleading. Nonconsequentialist theories are also concerned with consequences, though (primarily) with logical (or conceptual) consequences rather than with empirical ones. So, for example, while Kant certainly understands punishment to be justified as an institution by its tendency to control crime, he says nothing to suggest that that tendency is merely empirical. Indeed, he explicitly says the opposite: "The necessity of the public lawful coercion does not rest on a fact, but on an a priori Idea of Reason, for, even if we imagine them [those subject to positive law] to be ever so good and righteous before a public lawful state of society is established, individuals, nations, and states can never be certain that they are secure against violence from one another, because each will have his own right to do what seems just and good to him, entirely independently of the opinion of others."[17] The "necessity" of public lawful coercion (including the institution of punishment) is a (logical) consequence of a coordination problem unavoidable for rational agents much like us (beings who cannot know each other's judgments in advance).

For this reason, the other common way of distinguishing between theories of punishment, that is, distinguishing those that are "forward-looking" from those that are "backward-looking", must also be mistaken. True, a typical nonconsequentialist theory such as Kant's is (largely) backward-looking when justifying the punishment of an individual. But consequentialist theories have trouble justifying the punishment of an individual using exclusively forward-looking reasons. A system of criminal law that does not require a past crime as a precondition for punishment, and some relationship between the amount of punishment and the crime, is vulnerable either to the charge that the system is unjust in that respect or to the charge that it is not (in that respect at least) a system of criminal law at all. That vulnerability is why the classic mixed view ceded the justification of individual acts of punishment to retributivism.

More important now, though, is that even Kant's theory is not backward-looking when justifying the institution of punishment. A "public lawful state" is not a result of any past act (a causal consequence) but a condition existing as long as (enough) justice is being done. A public lawful state is (in part) a logical (or conceptual) consequence of just punishment; the punishing is (part of) what makes the public

[17] Kant (1999, p. 116).

state lawful. The institutional part of Kant's theory of punishment is more "sideways-looking" than either backward-looking or forward-looking.

Defining nonconsequentialist theories negatively, that is, as those theories not dependent on empirical consequences, may suggest that there are two distinct forms of nonconsequentialism: (a) those that merely avoid considering empirical consequences; and (b) those—the doubly nonconsequentialist—that avoid logical consequences as well. A doubly nonconsequentialist theory might, for example, simply point to a moral rule to establish a right to punish or even a duty to punish.

We may, I think, put aside the possibility of a doubly nonconsequentialist theory. Such a theory would beg the question of justification—or, at least, push it back, for example, to justifying the moral rule cited. The rule must, after all, be specific to punishment or it could not apply to punishment without a connecting argument, either (partly) empirical (and therefore consequentialist) or wholly logical (and therefore simply nonconsequentialist).

Given the ambiguity of the term "consequence", it seems better to describe theories typically called "consequentialist" as "empirical" (or "externalist")—and their competitors not as "nonconsequentialist" but as "conceptualist" (or "internalist").[18] Empirical theories depend crucially on statements that may be false in worlds very much like this one; conceptual theories do not. The statements on which conceptual theories depend are either logical truths strictly so called or (more often) statements that could only be true in worlds radically different from this one (for example, worlds in which people could not do one another serious harm). Though the distinction is not sharp in principle, in practice few, if any theories, lie on the fuzzy line it defines.

Benn's distinction is, however, not between consequentialist and nonconsequentialist theories but between "utilitarian" and "retributive". That distinction, though still in wide use today, is even more misleading than the consequentialist–nonconsequentialist distinction for at least two reasons. First, Benn's distinction is distinctly narrower. It does not allow for non-utilitarian theories that are also non-retributive (for example, nonconsequentialist education theories).[19] Second, Benn's distinction imposes on all non-empirical theory the burden of explaining in what sense punishment "returns" the crime to the criminal ("strikes back" at him, "evens the score", or the like). While some of Benn's retributivists, most notably Kant, do explicitly understand punishment as retributive in this sense, even a number of Benn's retributivists prefer other ways of talking about punishment, ways not

[18] I regard the choice between the terms "conceptual" and "internal" (or "empirical" and "external") as a matter of taste so long as the underlying distinction is clear. Analytically trained philosophers seem to prefer the empirical–conceptual distinction; those trained in the continental tradition, the internal–external.

[19] So, for example, Benn–Edwards (1967, p. 30)—is inclined to treat education theory as a disguised utilitarianism rather than as truly retributive. Education, after all, is a consequence. This is not a mere slip. Benn explains his underlying rationale in Benn (1958, pp. 326–327):

> The retributivist refusal to look to consequences for justification makes it impossible to answer this question within his terms. Appeals to authority apart, we can provide ultimate justification for rules and institutions, only by showing that they yield advantages. Consequently, what pass for retributivist justifications of punishment in general, can be shown to be either denials of the need to justify it, or mere reiterations of the principle to be justified, or disguised utilitarianism.

literally retributive. So, for example, Hegel speaks of punishment as annulling the wrong the person punished did. Philosophy should avoid categories that rule out (or even just seem to rule out) such live possibilities in advance (or invite rough treatment to make them fit).[20] In what follows, therefore, I sort theories into those that are empirical (externalist) and those that are conceptual (internalist). In principle, then, any claim about punishment in which the term "consequence" (or an instance of it, such as "education" or "deterrence") is central will have both an empirical version and a conceptual version (though one of the pair may be far less plausible than the other).

4 Empirical Theories

Most empirical theories of punishment are preventive, that is, they treat punishment as (primarily) a means of *controlling* objectionable behavior (by deterring, reforming, incapacitating, or otherwise keeping it from happening too much). Of course, the consequence that justifies punishment is not simply controlling the behavior to some degree, but reducing it enough to repay necessary costs. When the reduction in objectionable behavior is too low or the costs too high, punishment cannot, according to an empirical preventive theory, be justified (as act or at least as institution).

However, there could be empirical theories that are not preventive. One sort would rely on some past (or present) event (apart from crime) to justify punishment as an institution. So, for example, a Christian who justifies criminal punishment by claiming that God commanded the institution (by some historical act) is offering an empirical justification of punishment that is not preventive. God's command (recorded in the Bible long ago, a past event, or enlightening a particular believer at this moment of grace, a present event) is (we are supposing) a contingent fact. The same would be true of a theorist who argued that the institution of punishment is required under an actual constitution or historical social contract.[21] The same would, however, not be true of a Christian who justified punishment as arising from the divine nature or of a theorist who understood the social contract as a representation of necessary relations among rational agents. These two would be offering conceptual justifications of punishment.

These two non-preventive empirical theories are interesting only insofar as they call our attention to possibilities that other ways of sorting theories of punishment do not—or, at least, make harder to sort. These two non-preventive theories help us to see the value of sorting punishment theory this way rather than as Benn did.

A third sort of non-preventive empirical theory actually has defenders. It understands punishment as expressing some proposition concerned with justice ("This is how wrong the criminal act was") and claims that the (contingent)

[20] For others who find the term "retributive" misleading, see Cottingham (1979), Walker (1999).

[21] Actually, this is not quite right. The appeal to an actual constitution or social contract would directly answer question 2—but it would assume an answer to question 1 (for example, that the institution of punishment is morally permissible if, and perhaps only if, authorized by an actual constitution or social contract).

satisfaction to victim or public arising from that expression repays the cost of punishment (whatever its effect on crime). Feinberg's influential article, "The Expressive Function of Punishment", did not (as later writers sometimes suggest) propose this theory, though it did discuss theorists who did propose it. What the article explicitly proposed was revising the standard definition to acknowledge that punishment is in part a way of expressing a moral view. The revision seemed necessary because (it seems) Feinberg could not otherwise recognize a conceptual relationship between legal punishment and moral condemnation. Yet, for Feinberg, that relationship was important. It allowed him to distinguish between "punishment" and mere "penalty". Punishment includes moral condemnation while mere penalty does not.[22] This explicit connection of punishment with morality was, I think, what made Feinberg's article inspiring to many *internalists*—whom I discuss in the next section. Recent *externalists* seem to have dismissed empirical expressivism.[23]

Insofar as empirical theories propose to justify punishment on the assumption that penalties can (at reasonable cost) be tuned to achieve a certain degree of social control, they are vulnerable not only to empirical evidence (if the world does not work as they assume) but also to its absence (because the proof of an empirical theory is that following it has the justifying consequences).[24] By the 1970s, it was clear that the social sciences could not then, or in the foreseeable future, give empirical preventive theories much empirical support. The social sciences could not, that is, say what effect, if any, statutory penalties, rehabilitation, exemplary punishment, or even incapacitation would have on the crime rate (much less whether those effects would repay the cost). If even relatively crude tuning of penalties to empirical consequences is in practice impossible, empirical preventive theories cannot justify punishment as an institution, much less choosing any institution of punishment over others, or choosing one punishment over another, except in some counterfactual world in which we would know much more than we do know here. A theory of punishment should be more practical than that.

The social sciences failed preventive theory in this way in part because they failed to develop an adequate general theory of crime. Following Bentham, the social sciences assume that crime is a matter of incentives and disincentives. Though no doubt true in some degree (that is, insofar as potential criminals are rational), that assumption does not itself *quantify* the part that the disincentive of statutory penalty or actual punishment contributes to controlling crime (as opposed, say, to the disincentive of arrest, pre-trial detention, or solemnities of trial). The

[22] Feinberg (1965). As Feinberg himself recognized, discussion of the expressive function has a long history within utilitarianism, going back at least to Stephen (1890, p. 99). Yet, I have not found a single theorist who, following Feinberg, explicitly revised the standard definition of punishment. The law certainly has an expressive function, but it generally seems to carry it out in words, that is, statutes, regulations, sentences, comments from the bench, and so on. The question that Feinberg raised is whether, in addition to these obvious forms of expression (or communication), the *act* of punishing adds something interesting. Modern empiricists seem to think not. For an exception, see Gahringer (1960).

[23] For an example of why utilitarians seem to find expressionism unattractive, see Walker (1991, pp. 21–24) where, however, the empirical version of the education theory seems to melt into it.

[24] Utilitarian theories are also famously (but apparently not decisively) vulnerable on moral grounds insofar as crime control is different from justice. See, for example, Rawls (1955, pp. 9–13).

crime rate may as easily go up as down following an increase in statutory penalty (or go down and then return to its earlier rate or rise even higher).

The social sciences failed preventive theory in another way. They failed to provide an adequate surrogate for a general theory of crime, such as a sufficiently large number of low-level generalizations. Crime is relatively rare, even in the life of many repeat offenders. Crimes may, in large part, be random events, impossible to predict individually or to associate reliably with any interesting variable whatever. We may, of course, be sure that large men will be more often guilty of battery than women or small men, or that the poor will be more likely to rob the rich on the street than the other way around, or even that there will be less crime when police officers are present than when they are absent. We may be sure (for conceptual reasons) that the institution of punishment *tends* (all else equal) to reduce crime. However, all the low-level generalizations we have do not together, even when combined with conceptual truths, amount to a sufficiently rich understanding of crime to *quantify* the costs and benefits of punishment enough to defend any empirical preventive theory.[25] That we feel justified in punishing even though the social sciences give us little help with explaining why suggests that the justification is not empirical (or, rather, not empirically preventive).

To say that traditional empirical theories of punishment have been in steep decline for several decades is, of course, not to say that no philosopher defends them any more but that philosophic defenders are becoming fewer, increasingly defensive, and (most interestingly) compromised.[26] Consider, for example, a 2005 article by Anthony Ellis.[27] Its title, "A Deterrence Theory of Punishment", suggests a traditional preventive theory. In fact, what Ellis presents is a theory "mixed" in a way that would have surprised Benn. Self-defense (or, rather defense of innocents) is, according to Ellis, what gives the right to punish. Deterrence merely explains how the legal system can effectively exercise that right. (We defend ourselves by issuing plausible threats to punish.) Ellis concludes the paper putting off to another day the empirical question that has undercut traditional preventive theory:

> It is surely incredible that punishment should have *no* deterrent effect. As I have already indicated, even the most hardened, or reckless, criminal will normally be deterred to *some* extent by the threat of punishment; the question is whether the extent of the deterrence justifies the costs involved.[28]

That Ellis recognizes quantification as an important question shows that he remains an externalist. That he does not answer it underscores the fundamental problem that

[25] For a useful short discussion of these empirical failures, see Martinson (1978–1979). For a fuller, more nuanced, and more recent discussion of what the social sciences can do now—or might do later—see Barnes (1990).

[26] So, for example, while listings under "punishment" in the *Philosopher's Index* have increased five-fold during the last forty years, listings under "punishment and deterrence" have remained almost steady (1967–77, 4; 1977–1987, 2; 1987–1997, 7; 1997–2007, 4).The pattern is much the same for "punishment and reform" (1967–77, 6; 1977–1987, 3; 1987–1997, 7; 1997–2007, 11): The pattern is somewhat different for "punishment and prevention" (1967–77, 1; 1977–1987, 3; 1987–1997, 4; 1997–2007, 8).

[27] Ellis (2005).

[28] Ellis (2005, p. 225).

externalists face. Neither Ellis nor any other empirical preventivist has been able to answer that question. None knows whether punishment does deter enough to be justified by its deterrent effect.[29]

While not much work in traditional preventive theory has been done for almost two decades, empiricism has not died out. What has largely replaced traditional empirical theory are approaches at once new and peripheral to punishment theory proper. One of these approaches began in the 1970s as a social movement concerned with "victims' rights". Like all important social movements, it consists of several (more or less) inconsistent strands. We may distinguish three, only half (yes, 1½) of which are externalist.

One strand, an updated version of empirical reform theory, seems to favor the designation "restorative justice". Like the old reform theory, restorative justice draws heavily on psychology, sociology, and related sciences to develop prescriptions for "healing". And, like the old reform theory, it seems most at home applied to petty offenders, especially juveniles. The chief difference between this strand and the old reform theory is a new emphasis on including the victim in the therapy, for example, in sessions devoted to trying to "resolve the dispute" outside the courtroom. This therapy, though "justice-based" in a way that older forms of therapy, such as psychiatry, are not, is justice-based only in seeking to satisfy the (subjective) *sense* of justice of those directly involved. The purpose of the therapy is not justice as such but benefit to victim, family, and criminal, and perhaps others, the benefit arising from their feeling that justice has been done. This strand does not claim that victims have a moral (or other pre-legal) right to have justice done, merely that granting the relevant legal rights would be good for criminal and victim—and, therefore, for society generally. To the degree that restorative justice has a moral foundation, it seems to be the same as the old reform theory's, some sort of utilitarianism, and to suffer from the same lack of empirical support. It does, however, have the practical advantage of fitting people's sense of justice—and that, no doubt, has much to do with its rising popularity among lawyers, social workers, and judges.[30]

A second strand of the victims' rights movement belongs to the general libertarian trend of the last few decades. For it, institutionalizing victims' rights is part of a general strategy to privatize criminal justice (and just about everything else). Victims' rights are primarily *property* rights and punishment is, therefore, a sort of restitution (or, at least, should be restructured to be).

Just as libertarians generally divide into economic (utilitarian) defenders of private property (like F.A. von Hayek) and moral (natural rights) defenders (like Robert Nozick), so too do the libertarian defenders of victims' rights. The economic

[29] Ellis (2005, p. 225) does not himself appeal to the social sciences to answer that question but to opinion: "at least as far as the current levels of imprisonment are concerned, virtually no-one thinks this is now the case." He does not explain how we are to distinguish between the fads that overwhelm common sense now and then (which an externalist should ignore) and a true rational assessment of evidence (which they are bound to respect). He does not even ask how such a common opinion could be justified (if it actually exists).

[30] For some recent defenses of this strand of restorative justice, see Walker (2006), Bennett (2006), Obold-Eshleman (2004), Gavrielides (2005), Dzur and Wertheimer (2002).

version emphasizes the crime-control effect of victims' rights (basically, deterrence and reform). The moral version emphasizes, instead, the victims' moral right to restitution (without reference to social welfare). For the moral version of this strand, victims' rights are pre-legal. The moral version therefore seems to rely on a conceptual theory of punishment.[31] Since the economic version presupposes a way to measure costs and benefits of recognizing victims' rights, it suffers from the same lack of empirical support as other empirical theories.

For the third strand, victims' rights, though private and moral, are personal rather than property rights. This strand argues that the law should recognize certain rights in the victim because they are the natural product of the crime. The crime creates a "debt" owed the victim that only the criminal's *suffering* can repay—unless victim and criminal agree on some substitute. While there is nothing wrong with government being involved in criminal justice (according to this third strand), government must come in, if it does come in, only (or, at least, primarily) as an agent of the victim, charged with undoing the wrong done him. The victim's rights should include whatever is necessary to see that government helps him get justice. Insofar as this third strand relies on principles for satisfying the victim's right that are pre-legal, it seems to rely on a moral theory in which punishment (in the standard sense) has no essential part. The criminal (or, rather, the wrongdoer) would owe the victim the same response were there no social institutions whatever. This third strand is wholly conceptual, deriving its justification of punishment from ordinary moral right.

Thus, while (empirical) reform and deterrence may be reviving within the victims' rights movement, the importance of that revival for the empirical theory of punishment is at least put in doubt by the strong presence of conceptualist elements.[32] More interesting, I think, is that we can read much that is written on behalf of first strand (restorative justice) and (the empirical version of) the second strand as concerned not with punishment at all but with "law and order" more generally—apology, restitution, reconciliation, and so on being forms of non-punitive social control which, for convenience, are administered within the criminal law.

That brings us to the second new trend in empiricist discussions of punishment, enforcement, belonging to a tradition of legal scholarship commonly called "law and economics". Mark Reiff has recently turned work on enforcement into an empirical theory in which punishment plays a significant—but small—part largely independent of traditional empirical theories.[33] Reiff divides enforcement into two broad categories, "previolation enforcement" (which prevents violation of a rule) and "postviolation enforcement" (which responds to a violation). The purpose of previolation enforcement is to reduce the perceived risk of violation so much that those the rule protects may act with reasonable assurance that the violation in question will not occur. Among means of previolation enforcement are police patrols, regular audits, locks on doors, statutory penalties, and absence of broken

[31] For a good example of this second strand, see Benson (1998, pp. 227–259).

[32] See, for example, Barton (1999).

[33] Reiff (2005)

windows. Among means of postviolation enforcement are not only arrest, trial, and imprisonment, civil suit for damages, and administrative penalties (such as loss of a license), but also notoriety, shunning, private revenge, and retroactive self-assessment (such as spontaneous shame). In effect, enforceability is a feature of a society taking everything into account. Punishment (indeed, the criminal law as a whole) is only one means of enforcement. A rule may be enforceable (and enforced) even if there is no statutory penalty for violating it. For example, most rules of courtesy are enforced without any help from police, courts, or jails.

Enforcement has a connection with rational decision theory (at least insofar as we assume most people to be more or less rational even when violating a rule). But, according to Reiff, the connection is not what (empirical) preventive theories of punishment suggests. For empirical preventive theory, what is crucial is the effect of punishment on potential rule-violators. For enforceability, in contrast, what is crucial is public perception, that is, the impression of those who benefit from having the rule in force. Do they feel secure enough?

In many respects, Reiff is clearly following Bentham. It is therefore significant that he treats *prevention* as secondary to the public sense that laws are generally obeyed. Enforcement gives assurance that doing as the law says is reasonable. Enforcement is not so much about prevention as about the reasonableness of assuming a reasonably well-ordered society. Much social order can be achieved without a significant number of people being deterred, incapacitated, or reformed. For Reiff, it seems, traditional preventive theories could all fail and yet punishment could have an empirical justification. For enforcement, what is crucial is not what (at reasonable cost) prevents a rational agent from violating a rule (or even what prevents agents less than rational) but what the public in general, or some appropriate segment of it, sees as sufficient to maintain enough social order.

For Reiff, the contribution of deterrence to enforcement is primarily previolation. The threat to potential violators adds to *our* willingness to act in ways we wish everyone to act (farm the land, manufacture useful products, go on harmless outings, and otherwise contribute to overall well-being). The contribution of (what Reiff calls) "retribution" is, in contrast, to give a *sufficiently satisfying* expression to the public's (and victim's) desire for retaliation for the rule-violation itself (and whatever harm flowed from it). Punishment need not be a fully just response in order to be *sufficiently* satisfying. The punishment need only be severe enough to "maximize the chances that [combined with other consequences of the rule-violation] individual instances of social conflict will neither result in lengthy cycles of Tit for Tat nor otherwise undermine social cooperation generally".[34] For purposes of enforcement, punishment need not be a just return for the crime ("a robust restoration of the social order"), as victims' rights advocates (and some internalist theories of punishment) require; punishment need only be *not intolerably* unjust (a response perhaps far short of full justice). Tolerable injustice is achieved when the victim's (and his representatives') "desire to obtain the requisite degree of

[34] Reiff (2005, p. 143).

retributive justice and his desire not to expose himself to further injury, effort, and expense are in what might be called a state of uneasy equilibrium."[35]

This, of course, is an important departure from most empirical theories. For them, actual punishment is a pure cost to be avoided when not necessary to maintain deterrence, reform, or incapacitation. For Reiff, in contrast, actual punishment may be a good practice more or less independent of deterrence, reform, and incapacitation. Actual punishment may be part of establishing (or maintaining) the sense of social order that all enforcement aims at.

How then do we determine that, as a matter of fact, the means of enforcement are enough to justify the claim that a rule is enforced? We must, after all, be able to make such determinations to provide an empirical justification of enforcement. Otherwise, enforcement theory would automatically suffer the same fate as empirical theories of punishment. Reiff offers a double answer. From the previolation perspective, a rule is in force insofar as those benefiting from it believe they can reasonably act in reliance on it (that is, reasonably act on the assumption that violations will be too few to matter). From the postviolation perspective, a rule is in force insofar as the response to the violation is sufficient to maintain (or reestablish) the sense that those benefiting from the rule may reasonably rely on it without extra-legal or illegal retaliation of their own.

Since enforcement is primarily a matter of public perception (though a perception resting, in the long run, on facts such as the number of crimes), we can evaluate enforcement by, for example, surveying those who are supposed to benefit from the rules in question. We can evaluate enforcement even if we have no way to evaluate the contribution of punishment as such (prevention) to overall security. In this respect at least, Reiff's theory of enforcement may make unnecessary empirical punishment theory. We may be able to evaluate packages of practices which include punishment without being able to evaluate the contribution of punishment as such to reducing crime. Enforcement theory may, then, make another sort of mixed view possible—with enforcement justifying punishment as an institution (whatever its effect on crime) and various conceptual theories of punishment answering questions 2–6.

5 Conceptual Theories Generally and Moralistic Theories in Particular

In contrast to empirical theories, conceptual theories (whether or not strictly "retributive") do *not* seek to justify punishment by pointing to an *empirical* relation between punishment and contingent consequences (such as a certain crime rate). For conceptual theories, the relation between punishment and its justification is "internal". No conceptualist need deny that punishment has some general tendency to control crime or that the tendency is *a* reason to have some punishment system rather than none. The tendency of punishment to control crime (that is, to deter crimes to some degree) follows deductively from the assumed rationality of those subject to the institution of punishment. (It is, after all, rational, all else equal, to

[35] Reiff (2005, pp. 150–151).

avoid pain, death, loss of liberty, and other typical punishments.) One can recognize that without ceasing to be a rigorous conceptualist.[36] All conceptualists need deny is that the actual (or probable) *degree* of punishment's contribution to control crime matters for understanding why we should (or should not) have this institution, practice, or act of punishment rather than another (or none at all). The primary arguments for punishment (justifications a–d) are conceptual (or, at least, not empirical in our sense).[37]

Conceptual theories may be divided into moralistic and legalistic. Moralistic theories of punishment have four (main) divisions: desert, paternalist, defense, and expressive. *Desert* theory (in its pure form) takes it as a conceptual truth that wrongdoing deserves an unpleasant response, that is to say, punishment. Punishment is justified because (and insofar as) it is deserved.[38] *Paternalist* theory holds that all justified punishment, or at least all justified punishment of rational agents, must aim (at least in part) at a certain good for those punished. This good may be subjective (Duff's "penance") or objective (Nozick's "connection with correct values" or Hampton's "education").[39] *Defense* theory holds that our right (or duty) to punish is simply an instance of the right (or duty) of each of us to defend himself or an innocent third party. The typical institution of punishment is justified because (and only insofar as) it exercises the individual's right of defense.[40] *Expressive* theory, in contrast, understands punishment as (primarily) an "expressive act", not meant to benefit or defend anyone, but simply to say—by rough treatment—what is appropriate and true.[41]

All four varieties of moralistic theory are internalist (in the sense used here) because they rely entirely on relations among concepts to justify punishment. For desert theories, punishment is simply giving wrongdoers what they deserve—and doing that is right or good in itself. For paternalist theories, the justification of punishment lies in the way punishment treats the wrongdoer—for example, as a being capable of learning justice from the punishment appropriate to the wrong. The

[36] The works on Kant cited above seem to me to have missed this point (and so, to have supposed that Kant's theory of punishment compromises with "consequentialism".They are nonetheless right in seeing that Kant is willing to use "consequences" (a just social order) in justification of the institution of punishment, something generally missed. For a good discussion of Kant's theory of punishment that makes clear that the entire argument is conceptual, see Hill (1999).

[37] Why the parenthetical hedge? We must, I think, include among "concepts" those propositions—such as that other people feel pain as I do—which, though perhaps not a priori truths, are part of what "everyone knows" (or, at least, what we would all, at our rational best, acknowledge)—what used to be called "self-evident truths" but are better thought of as empirical propositions (more or less) constitutive of rationality (at least for people much like us). They are necessarily evident to rational agents but not necessary truths "in themselves". They may well be false in possible world distant from ours.

[38] For a good statements of the desert version of retributivism, see: Davis (1972), Kleinig (1973), Scheid (1997), Husak (1992), Moore (1993).

[39] Duff (1986), Nozick (1981, pp. 363–397), Hampton (1984). For a good critique of paternalist theories, see Schafer-Landau (1991).

[40] Montague (1995), Farrell (2004). For a critique of Montague (and, implicitly, Farrell), see Davis (1997).

[41] Among important discussions of the expressive theory are: Skillen (1980), Primoratz (1989), von Hirsch (1993), Metz (2000). For an internalist critique of expressive theory, see Davis (1991).

seriousness of the wrong determines what penalty is appropriate to teach the lesson that the crime shows the wrongdoer needs to learn. For defense theories, the justification of punishment lies in its being the exercise of the right of self-defense or defense of a third party, not in what it actually accomplishes (or even in what it is likely to accomplish). The right of defense limits punishment to what is necessary for defense.[42] For expressive theories, the justification of punishment lies in what the punishment "says" (and how well it says it). The penalty is a condemnation, denunciation, censure, or other reaffirming of the wrongness of the punished act. The expression should be as emphatic as the crime is morally bad; the more severe the punishment, the more emphatic the expression is.

While desert theories seem to be the direct descendants of traditional retributivism, paternalist theories superficially resemble traditional *reform* theories, defense theories superficially resemble incapacitation theories (insofar as defense of self or another innocent incapacitates the criminal), and expressive theories similarly resemble traditional *deterrence* theories (condemnation, denunciation, censure, and other forceful reassertion of the law resembling a deterrent threat). All four forms of the theory nonetheless differ fundamentally from any empirical theory. According to the paternalist theory (in its pure form, at least), punishment would be justified even if wrongdoers never repent or learn as a result of punishment. What is important—important because it respects the moral personality of the wrongdoer—is that the right punishment be imposed with the right intention. For defense theory (in its pure form), punishment would be justified even if it never succeeded in defending anyone. What is important is that we have the right (or duty) under certain conditions, and that we act with the right intention (to defend), not that we exercise that right (or duty) successfully (much less that we exercise it successfully enough). In much the same way, according to the expressive theory (in its pure form), punishment is justified even if the emphatic expression has no effect on the crime rate or even on the wrongdoer's later conduct or that of anyone else. Reaffirming the wrongness of an act may (all else equal) itself be not only right (morally permissible) but good.

All moralistic theories share the assumption that punishment belongs to ordinary morality (rather than to the law as such). Moralistic theories use ordinary moral practices (such as disciplining children) to understand punishment (with legal punishment only a special case). The institution of punishment is not itself a relevant consideration in the justification of punishment.

Moralistic theories differ primarily in the part of ordinary morality to which they look for justification. Desert theory treats punishment as (negative) rewarding (typically using prize-giving as the analogy). Paternalist theory treats punishment as correction or teaching. Defense theory treats punishment as a mere extension of ordinary self-defense (or defense of innocents). Expressive theory treats punishment as a gesture (a non-verbal moral statement like spitting in the criminal's face). Insofar as morality offers other analogues to punishment, other moralistic theories are possible, for example, ones relying on forfeiture, restitution, or satisfaction of a

[42] But see Alexander (1991), which argues that defense theory lacks the resources for a theory of proportion.

promise.[43] One great problem with moralistic theories is that there are so many kinds, all resting on plausible claims of the same sort (the analogy with some moral practice). Insofar as *each* claim to provide *the* justification of punishment, the variety itself must condemn the method all rely on.[44]

It is, I think, a mistake to consider any of these moralistic theories as offering the only justification of punishment. They are much more plausible if understood as offering *a* justification—consistent with others—establishing either that punishment is morally permissible or that it is morally good (in certain cases). In general, theories relying on a moral right (defense, forfeiture, restitution, and so on) will show that punishment (the institution) is morally permissible, while those that rely on good activities that presuppose a moral right derived in some other way (giving prizes, teaching, speaking the truth, and so on) seem designed to show that punishment is morally good (if morally permissible for other reasons). So, for example, while everyone can agree that giving people what they deserve is a good thing, we all also seem agreed that not everyone has the right to give someone what he deserves. Jane may have won the race, but no spectator has the right to give her the prize. Only a designated official has that right (and that right depends on something other than desert, for example, the rules of the contest all participants accepted). Anyone invoking desert to justify punishment (as morally good) will have to invoke some other theory to explain why "we" (the state, the people, or whoever) have the right of punishment.[45] Any theory relying on simple moral right will, of course, have trouble justifying the punishment of any "victimless crime" (that is, a crime without an identifiable victim or potential victim).

6 Legalistic Theories and the Future

Unlike moralistic theories, legalistic theories assume that (justified) punishment is a practice (largely) confined to (relatively just) legal systems.[46] Analogies with other moral practices can be of only limited use. This rejection of moral analogies may explain why there is today only one important form of legalistic conceptualism, "the fairness theory" (also known as "benefits-and-burdens," "reciprocity," "unfair

[43] Indeed, some of the possibilities given here have been realized. For moralistic retributivism relying on analogy with forfeiture, see Goldman (1979), Morris (1991), Simmons (1992, pp. 148–161), McDermott (2001). For moral restitution, see Barnett (1977), Cederblom (1995).

[44] Note, however, that at least some moralists only make the modest claim that they are providing a *partial* justification of punishment (rather than the only justification). See, for example, Morris (1981), Hampton (1992) in which she adds expressivism and—in n2—deterrence to her educational justification of punishment.

[45] Defense of innocents is an exception because there does seem to be a duty to defend an innocent (other than oneself). That duty is, however, limited to situations where the cost and risks to the defender are not significant. The criminal law probably does not satisfy that condition. The costs of criminal justice are high—and the risks to other innocents are, in some jurisdictions at least, substantial.

[46] For important discussions of the fairness theory (and related issues), see Dagger (1993) and (a critic) Ellis (1997).

advantage", or "restoration" theory). Though apparently a recent development, its origins are confused. Herbert Morris' 1968 article in the *Monist* is generally cited as the first to state the theory, yet he never discussed justification of punishment in that article (or even the aim, function, rationale, or point of punishment). He simply assumed a society having a certain cooperative *structure* and then appealed to that structure to help explain how a criminal could have a right *to be* punished.[47] When Morris finally did suggest a (partial) "justification" for the institution of punishment more than a decade later, what he suggested was a paternalistic theory, suggesting it without a hint that he had changed his mind.[48] His paternalist theory of punishment was, it seems, intended to explain no more than how the institution of punishment (and acts under it) could be morally good. He was not offering a complete theory of punishment.

John Finnis seems to have been the first to endorse the view that maintaining a fair balance between burdens and benefits is "the most specific and essential aim of punishment" as an institution (which is *almost* to say "the" justifying aim). Though Finnis did this in a well-known article in *Analysis* four years after Morris' in *The Monist*, no one seems to have given him the credit. Perhaps they did not because he himself claimed only to be clarifying ideas Jeffrie Murphy had presented a few months before.[49]

Except for Morris, Murphy is the only writer commonly cited as the first defender of the fairness theory. Yet, he is not. While he did in fact do much to develop the theory, he initially attributed it to Kant and soon argued that it could *not* apply in any society much like ours.[50] Having denied the theory any practical employment, he eventually decided that Kant held no theory of punishment whatever.[51] So, as far as Murphy is concerned, the fairness theory is a good-for-nothing of unknown ancestry.

A few writers have listed me among the theory's defenders.[52] But I am not (though I do like the theory—properly understood). My view, one held since I began writing about punishment, is that there is so much to be said in defense of the institution of punishment that debate over something properly called "the" justification of punishment simply obscures the obvious.[53] Different people may rely on one or more of several different moral rights to establish the moral permissibility of punishment. They may also show that punishment is morally good

[47] Morris (1968). This article was, of course, widely read an astonishingly powerful defense of retributivism in general or the fairness theory in particular. I admit to being so converted to retributivism when I read the article in the mid-1970s. Why the article should be so misread I cannot say.

[48] Morris (1981).

[49] Finis (1972), Murphy (1971a).

[50] Murphy stated the theory in Murphy (1971b) and then dismissed it as unrealistic in Murphy (1973).

[51] Murphy (1987).

[52] See, especially, Philips (1985), Moore (1989), von Hirsch (1991): 549–380, Schafer-Landau (1996), Ellis (1997)—though Ellis understands that my primary interest is sentencing and related matters.

[53] Davis (1983).

enough to practice by appeal to one or more of several different purposes, functions, aims, rationales, or points. They may even be able to rely on one or more arguments to show that the institution is morally or rationally required. There is no need for them to agree on the justification. For all that disagreement, they will agree on an institution having (roughly) the same structure (something meeting the standard definition). I have followed Morris in making certain assumptions about that structure as a way to reach the questions that interest me (3–6). But my theory of punishment (1), if I have one, is a mix of the fairness theory, some moralistic theories (a wholly internalist "mixed view"), and perhaps an internalist (or externalist) version of enforcement.

Though Finnis seems to be the first writer explicitly to adopt the fairness theory, he is not the only one to adopt it. Since the late 1970s, the theory has had a fair number of supporters (and many critics who treated it as an established position). Generally, the supporters seem to adopt the theory because it fits a moral (or political) theory that is their main focus.[54]

The fairness theory holds that legal punishment (and close analogues) are justified (that is, morally permissible, positively good, or both) insofar as it supports the (relatively just) distribution of benefits and burdens that a relatively just legal system (or similar practice) creates. A relatively just legal system is a cooperative practice from which each benefits if others generally do their part and in which doing one's part is sometimes burdensome. According to the fairness theory (in its pure form at least), the institution of legal punishment is justified if (and only insofar as) punishment keeps lawbreakers from gaining an unfair advantage over the law-abiding. Punishment, if just, necessarily takes back the unfair advantage the crime as such takes (or, at least, some fair equivalent of that advantage). Punishment, when justified, is justified as corrective justice, that is, as part of maintaining a just legal order. Maintaining a just legal order is good in itself and—all else equal—morally permissible.

Though the fairness theory has an obvious affinity to certain theories of distributive justice (especially, Rawlsian social contract), it presupposes no particular moral or political theory.[55] All it presupposes is that there can be an equivalence between crime and (just) punishment assuring that (in general at least) legal punishment of certain people in certain ways will (as a conceptual matter) help to maintain overall distributive justice (however defined). That presupposition has provoked much criticism. Much of the criticism, perhaps all of it, presupposes a mistaken standard of adequacy. Sometimes the mistake sets the standard too high; it presupposes that the fairness theory must show that the punishment *in fact* fits the crime, that is, that the fit is pre-legal, a determinate natural relationship between crime and punishment. The critic might then ask, for example, "How can any punishment, say, eleven years in prison, take back the unfair advantage of

[54] Among the chief supporters are: Gewirth (1978, pp. 294–299), Hoekema (1980), Sadurski (1985, pp. 221–258), Scher (1987, pp. 69–88).

[55] For an extended defense of this point, see Davis (1988).

manslaughter?"[56] Sometimes, however, the mistake is more subtle, a failure to take the institution of punishment (the criminal law) seriously enough.

Often, treating people fairly means treating them according to the rules of some *voluntary* cooperative practice. Few, if any, systems of criminal justice are voluntary (except for visitors or immigrants). We are born into them. We can escape only if they let us go or we become criminals by leaving without permission. Clearly, the fairness theory cannot be about fairness in this sense (without being implausible). There is, however, another sense of fairness, that is, treating people according to the rules of a cooperative practice they would, *at their rational best*, have entered voluntarily (even though, as a matter of fact, they had no choice). The "would" signals a counterfactual—or, rather, two of them. The first counterfactual concerns the context of choice, that is, what the other options are supposed to be (in a choice that never existed); the second, how the choosers must be changed to be at their "rational best" (though they may never in fact undergo that change). Some criticism of the fairness theory (such as Murphy's) includes among the options available a legal system embedded in an ideally just economic order.[57] Since the ideally just is always better than the actual, no actual institution of punishment can be justified in that context (at least insofar as justification is comparative). Other criticism of the fairness theory sets too low a standard for rationality (for example, mere economic self-interest with just the information people actually have). If such people would accept almost any system of criminal law short of a death camp (as Hobbes may have supposed), then the fairness theory justified in that way would justify almost any system of criminal law. The fairness theory would prove too much. A good theory of punishment should have more critical power than that (that is, it should at least help us develop a discriminating answer to question 2).

Though I do not endorse the fairness theory as "the" justification of punishment, I think it can work as "a" (partial) justification of punishment provided two conditions are met. First, the theory must claim to rely on a moral ideal (rather than a moral right) and therefore to be understood as explaining why it is (sometimes) good to punish rather than why it is morally permissible. While correcting injustice is certainly a moral ideal (something everyone recognizes as, all else equal, a good thing), not everyone has the right to correct a particular injustice. The right requires a distinct argument (say, an appeal to the right to defend innocents or a delegation of

[56] For a good example of this mistake, see Dolinko (1994, p. 508):

> Davis's market allows us, for example, to decide that voluntary manslaughter deserves a lesser punishment than murder, but we cannot say what punishment for voluntary manslaughter actually suffices to remove that crime's "unfair advantage." Thus, California recently raised its maximum penalty for voluntary manslaughter from six to eleven years, and Davis's market model leaves us unable to decide whether the state has at last authorized the proper, deserved punishment for this crime or has instead chosen a penalty nearly twice what the crime deserved.

It seems not to have occurred to Dolinko that justice (and therefore the constraints of a reasonable theory) might leave a range of choices in which, given other penalties, both six years and eleven years for manslaughter might be morally permissible and not ruled out by other considerations. It also seems not to have occurred to Dolinko that, given other penalties for other crimes (say, murder or mayhem), eleven years might be too high.

[57] Murphy (1973).

the right of self-government). Second, a plausible version of the fairness theory must presuppose, when answering question 1, a choice between the criminal law and competing methods of social control within *a society much like ours*. Economic rationality may provide a useful interpretation of "rational best", but only if the context in which that sort of rationality is exercised resembles the perfect market on which economic theory rests (competent adults, full information, enough time to deliberate, no externalities, and so on). Insofar as the context of choice departs from the perfect market, the rules of procedure should be adjusted to compensate (much as Rawls tried to do in his "original position"). Much the same would be true if another sort of rationality were assumed.

Within a fairness theory so structured, the claim that punishment takes back the unfair advantage criminals take by crime would concern logical relations between crimes, penalties, and the rest of the legal order. The equivalence of a certain penalty (eleven years in prison) and a certain crime (manslaughter) would be much like the equivalence of a certain piece of paper (a $100 bill) and some commodity or service (a cell phone or oil change). They are equivalent only within, and only because of, the structure of the society in question (respectively, a market or legal order). While a free market imposes some constraints, the constraints seldom, if ever, yield a unique "true price". The same commodity or service might, on a different day or with a different buyer or seller, sell for somewhat more or somewhat less—without the price being too high or too low to be fair (in the counterfactual sense). The relation that eleven years in prison has to manslaughter might be the same. Eleven years might be a fair equivalent for manslaughter even though five years in prison would be too.

What next? For now, we now seem to have enough theories of punishment. What we need is more care in sorting them according to the questions they answer (especially, the sort of justification they offer)—and the method they use to answer those questions (empirical or conceptual). The achievements of the last half century were, it seems, made possible by the 1950's development of better tools for approaching punishment theory. We now seem to need another improvement in tools rather than yet another theory of punishment. Among those improvements would be a (more or less) canonical statement of explicit criteria of adequacy. I offer the following as a starting point: At a minimum, a theory of punishment should explain:

1) Why (and under what conditions) establishing (or maintaining) an institution of punishment is morally permissible;
2) Why (and under what conditions) we should maintain such an institution (including, perhaps, a showing that it is morally or rationally necessary for us to maintain it);
3) What implications, if any, the theory has for practice (for example, does it require us to have information we do not have, to conduct ourselves in ways we consider clearly wrong, or to revise the way criminal justice now operates); and
4) How (and under what conditions) the theory helps us to resolve questions 2–6.

These four criteria seem to me implicit in the half century of punishment theory I have just surveyed. What other criteria should be added?

The four criteria assume the standard definition of punishment (rather than requiring one) and therefore also assume that criminal acts are acts that violate the criminal law. That seems to me reasonable. Why go back to trying to explain the difference between "divine punishment" and "human punishment"? The four criteria say nothing about what should be made criminal because guiding the making of the criminal law belongs to the theory of legislation or enforcement. Philosophers of punishment (as such) generally have little to say on the subject. The four criteria also have nothing about whether "the state" (or government) should take responsibility (even in part) for organizing, operating, or funding an institution of punishment because the role of government also seems a question of political philosophy rather than punishment theory proper. If we are to have an institution of punishment, there are reasons to have government administer it—moral as well as practical—though in some places it has seemed reasonable for private groups, such as the "vigilance committees" of 1840s Colorado or the "bonds" of medieval Scotland, to do it. The problems of understanding punishment do not seem to change much because of who (or what) administers the institution. That seems a good reason not to require a theory of punishment to explain the role of government in punishment.[58]

Acknowledgments An early version of this paper was presented to the Humanities Colloquium, Illinois Institute of Technology, Chicago, on February 8, 2008. I should like to thanks those present—as well as Angelo Corlett, Brian Ellis, Steve Kershnar, and Don Scheid—for their helpful comments.

References

Alexander, Larry. 1991. Self-defense, punishment, and proportionality. *Law and Philosophy: An International Journal for Jurisprudence and Legal Philosophy* 10 (August): 323–328.
Armstrong, K.G. 1961. The retributivist hits back. *Mind* 70 (October): 471–490.
Barnes, J. 1990. *Models and interpretations*. Avon, UK: Cambridge University Press.
Barnett, Randy E. 1977. Restitution: A new paradigm of criminal justice. *Ethics* 87 (July): 287–301.
Barton, C.K.B. 1999. *Getting even: Revenge as a form of justice*. Chicago: University of Chicago Press.

[58] There are, then, important differences between my four criteria and those that Murphy (1987, pp. 510–512) offers. According to Murphy, a "complete" theory of punishment should answer all five of the following questions:

1. What is the nature of crime and punishment?
2. What is the moral justification of punishment?
3. What is the political justification of punishment?
4. What are the proper principles of criminal liability?
5. What are the appropriate punishments?

Two of Murphy's five questions are beyond punishment theory proper: one (1), because the answer (the standard definition) is now assumed; and the other (3), because its answer belongs to political theory (as Murphy's nomenclature acknowledges). The other three questions cover much the same ground as my four criteria, but are less precise (ignoring, for example, the distinction between the four senses of "justification"). They are also more demanding. I do not require a theory to answer his questions 4 and 5, merely to "help" answer them. Indeed, Murphy's five questions are so demanding that I cannot think of a single writer on punishment whose theory would count as "complete" under it. No wonder Murphy concluded that Kant did not have a theory of punishment!

Benn, Stanley I. 1958. An approach to the problems of punishment. *Philosophy* 33 (October): 325–334.
Bennett, Christopher. 2006. Taking the sincerity out of saying sorry: Restorative justice as ritual. *Journal of Applied Philosophy* 23 (May): 127–143.
Benson, B.L. 1998. *To serve and protect: Privatization and community in criminal justice*. New York: New York University Press.
Brooks, Thom. 2003. Kant's theory of punishment. *Utilitas: A Journal of Utilitarian Studies* 15 (July): 206–224.
Cain, James. 2002. On the problem of hell. *Religious Studies* 38 (September): 355–362.
Cederblom, J. 1995. The retributive liability theory of punishment. *Public Affairs Quarterly* 25 (October): 305–315.
Corlett, J. Angelo. 2006. *Responsibility and punishment*, 3d ed. Dordrecht, The Netherlands: Springer.
Cottingham, John. 1979. Varieties of retribution. *Philosophical Quarterly* 29 (July): 236–246.
Dagger, Richard. 1993. Playing fair with punishment. *Ethics* 103 (April): 473–488.
Davis, L.H. 1972. They deserve to suffer. *Analysis* 32 (March): 136–140.
Davis, Michael. 1983. How to make the punishment fit the crime. *Ethics* 93 (July): 726-752.
Davis, Michael. 1988. The relative independence of punishment theory. *Law and Philosophy* 7 (December): 321-350.
Davis, Michael. 1990. Recent work in punishment theory. *Public Affairs Quarterly* 4 (July): 217-232.
Davis, Michael. 1991. Punishment as language: misleading analogy for desert theorists. *Law and Philosophy* 10 (April): 310–322.
Davis, Michael. 1997. Review—Phillip Montague, punishment as societal defense. *Ethics* 107 (April): 532–534.
Dolinko, David. 1991. Some thoughts about retributivism. *Ethics* 101 (April): 537–559.
Dolinko, David. 1994. Mismeasuring 'unfair advantage': A response to Michael Davis. *Law and Philosophy* 13 (November): 493–524.
Duff, R.A. 1986. *Trials and punishments*. Cambridge: Cambridge University Press.
Dzur, Albert W., and Alan Wertheimer. 2002. Forgiveness and public deliberation: The practice of restorative justice. *Criminal Justice Ethics* 21 (Winter-Spring): 3–20.
Edwards, Paul, ed. 1967. *The encyclopedia of philosophy*, vol. 7. New York: Macmillan.
Ellis, Anthony. 1997. Punishment and the principle of fair play. *Utilitas: A Journal of Utilitarian Studies* 9 (March): 82–97.
Ellis, Anthony. 2005. A deterrence theory of punishment. In *Philosophy and its public role: Essays in ethics, politics, society and culture*, ed. William Aiken, John Haldane, 210–226. Exeter, UK: Imprint Academic.
Farrell, Daniel M. 2004. Capital punishment and societal self-defence. In *Philosophy and its public role: Essays in ethics, politics, society and culture*, ed. William Aiken, 241–256. Exeter, UK: Imprint Academic.
Feinberg, Joel. 1965. The expressive function of punishment. *Monist* 49 (July): 397–423.
Finis, John. 1972. The restoration of retribution. *Analysis* 32 (March): 131–135.
Flew, A.G.N. 1954. The justification of punishment. *Philosophy* 29 (October): 291–307.
Gahringer, Robert E. 1960. Punishment as language. *Ethics* 71 (October): 46–48.
Gavrielides, Theo. 2005. Some meta-theoretical questions for restorative justice. *Ratio Juris: An International Journal of Jurisprudence and Philosophy of Law* 18 (March), 84–106.
Gewirth, Alan. 1978. *Reason and morality*. Chicago: University of Chicago Press.
Goldman, Alan. 1979. The paradox of punishment. *Philosophy & Public Affairs* 9 (Fall): 42–58.
Hampton, Jean. 1984. The moral education theory of punishment. *Philosophy and Public Affairs* 13 (Summer): 208–238.
Hampton, Jean. 1992. Correcting harms versus righting wrongs: The goal of retribution. *UCLA Law Review* 39 (August): 1659–1702.
Hart, H.L.A. 1960. Prolegomenon to the principles of punishment. *Proceedings of the Aristotelian Society* 60: 1–26.
Hegel, G.W.F. 2008. *The philosophy of right*. www.marxists.org/reference/archive/hegel/prindex.htm.
Hill, Thomas. 1999. Kant on wrongdoing, desert, and punishment. *Law and Philosophy* 18 (July): 407–441.
Hoekema, David. 1980. The right to punish and the right to be punished. In *John Rawls' theory of social justice: An introduction*, ed. H.G. Blocker, and E.H. Smith, 239–269. Athens, Ohio: Ohio University Press.
Husak, Douglas. 1992. Why punish the deserving?. *Nous* 26 (December): 447–464.

Kabay, Paul. 2005. Is the status principle beyond salvation? Toward redeeming an unpopular theory of hell. *Sophia* 44(May): 91–103.
Kant, Immanuel. 1999. Metaphysical elements of justice, 2d ed. (trans and ed: Ladd, John). Indianapolis, IN: Hackett.
Kleinig, John. 1973. *Punishment and desert*. The Hague: Martinus Nijhoff.
Martinson, Robert. 1978–1979. New findings, new views: A note of caution regarding sentencing reform. *Hofstra Law Review* 7: 243–258.
McCloskey, H.J. 1962. The complexity of the concept of punishment. *Philosophy* 37 (October): 307–325.
McDermott, Daniel. 2001. The permissibility of punishment. *Law and Philosophy* 20 (July): 403–432.
McPherson, Thomas. 1967. Punishment: Definition and justification. *Analysis* 28 (October): 21–27.
Merle, Jean-Christophe. 2000. A Kantian critique of Kant's theory of punishment. *Law and Philosophy: An International Journal for Jurisprudence and Legal Philosophy* 19 (May): 311–338.
Metz, Thaddeus. 2000. Censure theory and intuitions about punishment. *Law and Philosophy* 19 (July): 491–512.
Montague, Phillip. 1995. *Punishment as societal defense*. Lanham, Maryland: Rowman & Littlefield.
Moore, Kathleen. 1989. *Pardons*. New York: Oxford University Press.
Moore, Michael S. 1993. Justifying retributivism. *Israel Law Review* 27 (Winter-Spring): 15–49.
Morris, Herbert. 1968. Persons and punishment. *The Monist* 52 (October): 475–501.
Morris, Herbert. 1981. A paternalistic theory of punishment. *American Philosophical Quarterly* 18 (October): 263–271.
Morris, Christopher W. 1991. Punishment and the loss of moral standing. *Canadian Journal of Philosophy* 21 (March): 53–79.
Murphy, Jeffrie G. 1971a. Three mistakes about retributivism. *Analysis* 31 (April): 166–169.
Murphy, Jeffrie G. 1971b. Kant's theory of criminal punishment. In *Proceedings of the third international Kant congress*, ed. L.W. Beck, 434–441. Dordrect: D. Reidel.
Murphy, Jeffrie G. 1973. Marxism and retribution. *Philosophy and Public Affairs* 2 (Spring): 217–243.
Murphy, Jeffrie G. 1987. Does Kant have a theory of punishment. *Columbia Law Review* 87 (April): 509–532.
Nozick, Robert. 1981. *Philosophical explanations*. Cambridge, MA: Harvard University Press.
Obold-Eshleman, Christa. 2004. Victims' rights and the danger of domestication of the restorative justice paradigm. *Notre Dame Journal of Law, Ethics and Public Policy* 18 (2): 571–603.
Philips, Michael. 1985. The justification of punishment and the justification of political authority. *Law and Philosophy* 5 (December): 393–416.
Primoratz, Igor. 1989. Punishment as language. *Philosophy* 64 (April): 187–205.
Quinton, A.M. 1954. On punishment. *Analysis* 14 (June): 133–143.
Rawls, John. 1955. Two concepts of rules. *Philosophical Review* 64 (January): 3–32.
Reiff, Mark R. 2005. *Punishment, compensation, and law: A theory of enforceability*. Cambridge: Cambridge University Press.
Sadurski, Wojciech. 1985. *Giving desert its due*. Dordrecht: D. Reidel.
Schafer-Landau, Russ. 1991. Can punishment morally educate?. *Law and Philosophy* 10 (May): 189–219.
Schafer-Landau, Russ. 1996. The failure of retributivism. *Philosophical Studies* 82 (July): 289–316.
Scheid, Don E. 1983. Kant's retributivism. *Ethics: An International Journal of Social, Political, and Legal Philosophy* 93 (January): 262–282.
Scheid, Don E. 1997. Constructing a theory of punishment, desert, and distribution of penalties. *Canadian Journal of Law and Jurisprudence* 10 (July): 441–506.
Scher, George. 1987. *Desert*. Princeton, New Jersey: Princeton University Press.
Simmons, A. John. 1992. *The Lockean theory of rights*. Princeton, New Jersey: Princeton University Press.
Skillen, J. 1980. How to say things with walls. *Philosophy* 55 (October): 509–523.
Stephen, J.F. 1890. *A history of the criminal law of England*. London: Macmillan.
Tunick, Mark. 1996. Is Kant a retributivist?. *History of Political Thought*. 17 (Spring): 60–78.
von Hirsch, Andrew. 1991. Proportionality in the philosophy of punishment: From 'Why Punish?' to 'How Much?'. *Israel Law Journal* 25 (Summer–Autumn): 549–380.
von Hirsch, Andrew. 1993. *Censure and sanctions*. Oxford: Oxford University Press.
Walker, Nigel. 1991. *Why punish?*. Oxford: Oxford University Press.
Walker, Nigel. 1999. Even more varieties of retribution. *Philosophy* 74 (October): 595–605.

Walker, Margaret Urban. 2006. Restorative justice and reparations. *Journal of Social Philosophy* 37 (Fall): 377–395.
Walls, Jerry L. 2004. A hell of a dilemma: Rejoinder to Talbott. *Religious Studies* 40 (June): 225–227.
Zaibert, Leo. 2006. Punishment as revenge. *Law and Philosophy* 25 (January): 81–118.