Published in On-Line Philosophy Conference
http://garnet.acns.fsu.edu/~tan02/OPC%20Week%20Four/
Accessed July 5, 2007 ("published" 2006)

Methodological Missteps: A Response to Brooks' "On Retributivism"

Michael Davis
Illinois Institute of Technology

Thom Brooks claims his paper has only one thesis—that, for epistemological reasons (in particular, requiring a "gold standard of desert"), "retributivism is impossible to enact as a practice". I think the paper has at least two other theses as well, both unacknowledged and more or less independent of the first. One is that there is only one true retributivism ("pure retributivism" or "retributivism strictly speaking"). This claim seems to rest on an unjustified, and (I believe) unjustifiable, Platonism. The second unacknowledged thesis is that this one true retributivism suffers (and must suffer) from certain flaws, in particular: 1) *moral rigorism* (forbidding the criminal justice system to show mercy or to deviate in any other way from what the criminal deserves for his crime);[1] 2) *methodological individualism* (requiring that desert be "a particular criminal's desert" unaffected by "other factors, such as society's equilibrium");[2] and 3) *methodological absolutism* (an inability to "choose punishments on account of how they might be related to each other and various crimes").[3]

What I propose to do here is, first, briefly dispose of Brooks' epistemological claim, then (at greater length) explain why retributivism is best thought of as a family of loosely related theories no one of which has the privilege of being "true retributivism" (even though some are certainly historically or conceptually closer to the core of retributive thinking than others) and, last, why, so understood, retributivism does not suffer from any of the three flaws Brooks claims true retributivism must suffer from.

The Gold Standard

Brooks concludes his paper with the following sentence: "If all judgments of desert and proportionality are *our* judgments that may differ from what is deserved in itself, then retributivism as such is impractical, if not impossible, to implement."[4] Our

judgments are, of course, all the judgments we have. There are no judgments without "us" (no infallible judge against whose judgment we can check our own). And—also of course—our judgments, including our judgments of desert, may be mistaken (inconsistent with "what is deserved in itself"). We are all residents of Plato's Cave and must make due with its dim and flickering light.

Brooks is doubtless right about this. The question is why this truth should make retributivism impractical (that is, impossible to implement). Consider the corresponding truth about any consequentialist theory of punishment. All judgments of consequences are *our* judgments; they *may* not be consistent with what the consequences actually are. Therefore any consequentialist theory of punishment is impractical. We have no way to know (with certainty) that we have done what the theory asks of us. We may argue in much the same way if, as utilitarians often do, we substitute expected utility (utility adjusted for probability) for simple consequences in the justification for punishment. Though adjustment for probability may seem to take account of the life's uncertainties, all judgments of expected utility must be our judgments; those judgments may not be consistent with expected utility in itself—consistent, that is, with what a calculation free of error would deliver. We cannot guarantee freedom from error. Therefore, even using expected utility to determine punishment is impractical. We can never know (with certainty) that we are acting as we should.

Clearly, there is something fundamentally wrong with this way of arguing. We should not demand certainty where certainty is impractical (and certainly not where it is altogether impossible). Certainty (understood as no chance of error) is not possible in any activity in which humans are involved. Hence, no theory of punishment should demand certainty in application of its principles. Methodological rationality requires any theory to settle for something less—just as we settle for something less in a mathematical proof—the satisfaction of a reasonable standard of care. If we later discover an error, we should—as reasonable creatures—make the necessary correction. But until we discover error, the proof stands, though we cannot rule out an error yet undiscovered. We can say we know the answer (even if we should not say we know it "with certainty").

Might Brooks respond that retributivism differs from mathematics (and almost all other human undertakings) in requiring certainty as a condition of application? He

might—and, for consistency's sake, should—make some such response. He would, of course, be wrong as a matter of fact. Retributive theories never require more than "reasonable certainty" for a judgment of deserved punishment. However, what Brooks says (in the sentence immediately preceding the one just quoted) is more plausible (but not so useful in refuting retributivism): "retributivism, strictly speaking, demands we make judgments about issues such as desert and proportionality where we cannot guarantee our accuracy."[5] Brooks thus admits the obvious (or, at least, seems to): retributivism (like any other plausible theory of punishment) recognizes human fallibility. In that respect, then, retributivism is as practical as any other theory of punishment. Brooks has no argument for the claim that retributive theory must adopt a standard of application impossible to satisfy. Brooks' epistemic conclusion cannot follow from the premise that "retributivism strictly speaking" presupposes that we know what a criminal deserves. We may never know "for certain" that our judgments of criminal desert are right, but we can nonetheless know them to be right in a somewhat weaker sense, the ordinary sense in which we are said to know things, what is captured roughly in the phrase "justified true belief". Though we cannot know (with absolute certainty) that our judgments are true, we can be justified in our beliefs; those beliefs may in fact be true; we are entitled to believe they are true until we have good reason to think otherwise; and we may justifiably act accordingly. Because we may know (in this ordinary sense) what a criminal deserves ("to a moral certainty" or "beyond a reasonable doubt"), Brooks' epistemic objection to retributivism need not concern us further. We may proceed to his other objections.

Retributivism Strictly Speaking

Most words ending with "ism" seem to have originated with an enemy of the belief in question. Those who first embrace a belief have no reason to give it a special name; they simply embrace it along with all the others they hold as "the simple truth". It is those who wish to criticize the belief who must separate it from the rest of what is believed, work out its implications, and give it a name to facilitate reference. An enemy of a belief is unlikely to be as charitable in its interpretation as a believer, as sensitive to

its nuances, or even as aware of its resources.[6] We must, then, be careful not to be too strict about what any "ism" means until we have checked with its friends. Few retributivists would, I think, agree with Brooks' characterization of "retributivism strictly speaking". Certainly, I do not, though I consider myself a retributivist (and, judging by whom Brooks list among his sources for what retributivism is, Brooks seems to agree). Any disagreement about what retributivism is cannot be settled by examining a Platonic form labeled "retributivism". There is no such form (and, if there were, we creatures of the Cave would have no way to know we were examining it). What we have instead is a human attempt to put into words certain characteristics of a set of theories, a construction rather than a discovery.

I shall now describe retributive theories briefly but in enough detail to make clear how they seem to be connected. The connections are, I think, plainly not rich enough to constitute a theory of punishment, "true retributivism". The connections yield instead a family of loosely related theories too varied for any plausible criticism to address all. Yet, many of theories have at least as good a claim to count as "true retributivism" as Brooks' candidate.

Having denied Brooks his unified target in this section, I shall show that retributivism in general need not suffer from rigorism and that at least one version of retributivism, my own, also avoids the other two flaws Brooks claims all true retributive theories must suffer (methodological individualism and methodological absolutism). I shall leave others the work of making the same point about their favored version of retributivism. I will have done all I need to do here. Brooks cannot have refuted "retributivism strictly speaking" if even one genuinely retributive theory is immune to his refutation.

What all retributive theories seem to share is the claim that the relation between crime and punishment is (primarily) conceptual (or "internal"). The justification of punishment is that punishment in itself is an appropriate response to crime. Retributive theories so understood would be uninteresting were there not another family of theories, the consequentialist (or utilitarian), that claim the opposite, that is, that the relation between crime and punishment is (primarily) empirical (or "external"), a matter of particular facts about how it happens to work (or about how we may, as a matter of fact,

make the world work for our benefit). For consequentialist theories, the justification of punishment is that it has certain (contingent) effects; it deters, reforms, incapacitates, or the like (or, at least, is likely to). Where these good effects are not sufficiently large, punishment cannot be justified. For a consequentialist, there is nothing inherent in the idea of crime (or wrongdoing) that could justify punishment.

That, I think, is about all retributive theories share. They divide into two importantly different kinds: moralistic and legalistic. Moralistic retributivism has three (main) subdivisions: desert, paternalist, and condemnatory theories. *Desert* theories take it as (more or less) brute fact about our concepts that wrongdoing deserves an unpleasant response, that is to say, punishment. Punishment is justified because it is deserved. Nothing more need be said (except how desert is to be measured). Giving people what they deserve is good in itself.[7] *Paternalist* theories hold that all justified punishment, or at least all justified punishment of rational agents, must aim at a certain good for those punished (or it is not punishment). This good may be subjective (Duff's "penance") or objective (Nozick's "connection with correct values").[8] *Condemnatory* theories, in contrast, understand punishment as (primarily) an "expressive act" not meant to benefit anyone.[9]

These three varieties of moralistic theory are retributive (in the sense used here) because all seek to achieve a good that is conceptually related to the crime. For desert theories, that good is simply giving wrongdoers what they deserve. Degree of desert determines severity of punishment. 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 crime. The seriousness of the wrong determines what penalty is appropriate to teach the lesson that the crime shows the wrongdoer needs to learn. For condemnatory theories, the justification of punishment lies in what the punishment "expresses". The denunciation should be as emphatic as the crime was bad; the more severe the punishment, the more emphatic the denunciation of the crime. The punishment tells the truth about the crime. For both paternalistic and condemnatory theories, "desert" (giving what is due) is defined (in part) in terms of some other good.

While desert theories seem to be the direct descendants of traditional retributivism, paternalist theories superficially resemble traditional *reform* theories, and condemnatory theories similarly resemble traditional *deterrence* theories (denunciation resembling a

deterrent threat). Both nonetheless differ fundamentally from the corresponding consequentialist theories. According to paternalist theories, 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. In much the same way, according to condemnatory theories, punishment is justified even if emphatic denunciation has no effect on the crime rate or on the individual's later conduct. Reaffirming the wrongness of the crime is good in itself, good enough (all else equal) to justify the punishment. Telling the truth about a crime is itself an important good

All *moralistic* theories share the assumption that punishment belongs to ordinary morality (rather than to the law in particular). Moralistic theories use ordinary moral practices (such as disciplining children) to understand punishment. For moralistic theories, legal punishment is only a special case. Moralistic theories differ from one another primarily in the part of ordinary morality to which they assign punishment. Desert theories interpret punishment as (negative) rewarding; paternalist theories interpret punishment as correction or teaching; and condemnatory theories interpret punishment as (implicit) moral statement. Other moralistic theories are possible, for example, one interpreting punishment as a form of self-defense or forfeiture.[10]

Legalistic theories, in contrast, assume that (justified) punishment is a practice (largely) confined to (relatively just) legal systems. The only important form of legalistic retributivism today is "fairness theories" (also known as "benefits-and-burdens," "reciprocity," or "unfair advantage" theories). All forms of the fairness theory hold that legal punishment (or any close analogue) is justified insofar as—as a conceptual matter—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 to be thought of as a cooperative enterprise from which each benefits if others generally do their part and in which doing one's part will sometimes be burdensome. According to fairness theories, the institution of legal punishment is justified if (it is a conceptual truth that) 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). Though fairness theories have an obvious affinity

with certain theories of distributive justice (especially, Rawlsian social contract), they presuppose no particular theory of distributive justice. All they presuppose 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) increase (or at least help to maintain) overall distributive justice (however defined).[11] Explaining that presupposition has proved difficult. Much of Brooks' criticism is in fact aimed at the difficulties. We may now consider the three particular criticism with which we began.

Rigorism

Brooks' characterization of "true retributivism" begins reasonably enough. "Retributivism", he says, "justifies punishing only persons who *deserve* punishment."[12] But he then adds, "This entails, first, that all persons who deserve punishment *must* be punished." Hence, according to Brooks, moral rigorism is an immediate consequence of retributivism's reliance on desert for its justification of punishment. The entailment in question presupposes an unstated premise, something like: *We must always give people what they deserve for their crime*. Yet, on any ordinary understanding of desert, the most we can say is that, *all else equal*, we *should* always give people what they deserve for their crime (because, all else equal, we should always give people what they deserve). If, however, someone deserves inconsistent things, say, life in prison for a crime, but mercy for repentance, then, all things considered, we are entitled to reduce the penalty to take account of the repentance. Retributivism sets an upper limit on what punishment can be—and, all else equal, a lower limit. But when all else is not equal, morality as a whole may take precedence over that part of morality concerned with punishment as such, canceling whatever rigor it might otherwise have. Non-criminal desert may reduce what a criminal should receive in punishment.

Of course, once we recognize any obligation to punish as merely *prima facie*, an obligation we have all else equal, we can see how other considerations beside non-criminal desert may justify punishing less than the crime deserves. I may, for example, justifiably fail to give you a gift you deserve if I cannot afford it, if giving it would harm

an innocent person, or if it would do some other serious mischief. You remain as deserving as ever, not just *prima facie* but all things considered, but other considerations beside your desert are relevant to what I should (or must) do about giving you what you deserve. Clemency is a problem only for retributivists who, like Kant, fail to distinguish between *prima facie* and actual obligation. There is nothing in "desert" as such to require us to deny that distinction—or to require us to suffer the difficulties to which denial of that distinction seems to lead.[13]

Yet, even if we were to deny the distinction between *prima facie* and actual obligation, retributivism need not suffer from rigorism. When it is said that a theory of punishment provides a justification of punishment, the term "justification of punishment" is ambiguous. There is at least one reasonable sense of justification consistent with avoiding rigorism even without appealing to the distinction between *prima facie* and actual obligation.

We may distinguish at least four (related) senses in which punishment may be "justified". First, it may be shown to be morally *permissible*. Justification of this sort treats morality as a "side constraint". Many acts, for example, wounding another in self-defense, are morally justified in this weak sense. They are (all things considered) morally alright, though some of the alternatives are morally alright too (say, risking one's own safety to avoid harm to an attacker). Second, acts may be morally justified in a stronger sense. Not only is doing them morally alright, but doing anything else instead is morally wrong. These acts are morally *required*. (So, for example, all else equal, keeping a promise is justified in this stronger sense; failing to keep the promise is morally wrong.) Third, some acts, whether or not *morally* justified (in either the weak or strong sense), may still be justified according to some non-moral standard (that is, permitted or required by it). And, fourth, some justifications seem to combine the weaker moral justification with a non-moral justification. For example, I may justify punishing you by pointing out that what I am doing is *both* legally required *and* morally permissible.

Brooks may be assuming that a theory of punishment must justify punishment in the morally strong sense. Without the distinction between prima facie and actual obligation, a theory of punishment would then necessarily require us to punish. My own theory—and perhaps many other retributivist theories as well—are not so demanding.

They instead seek to show that punishment is, subject to certain constraints, morally permissible. Among those constraints is, I believe, a prohibition of inhumane penalties. Whether a society should have an institution of punishment and, if it should, what wrongdoing that institution should punish, are distinct questions of policy a retributive theory need not answer. Morality does not require us to do what it merely permits. Even Kant would agree to that.

Methodological Individualism and Absolutism

Having (rightly) noted the central place of desert in retributivism (while misconstruing what it entails), Brooks makes two other points with which I agree: a) that, according to retributivism, criminals "deserve a punishment that is equal in value to the gravity of their offense"; and b) that that value "derives solely from the criminal's own act".[14] But Brooks and I again part company at his next five words, "and not from anywhere else". Of course, one way to read these five words is as no more than a commonplace of retributivism. What derives "solely" from the criminal's act cannot, as a conceptual matter, derive from anything that merely happens to be connected with the act. There are, I should add, problems even with this commonplace. Most retributivists have thought that one of its consequences is that an attempt should be punished as severely as the corresponding complete crime. So, for example, a would-be murderer should not (it is said) receive less in punishment because he missed his target's heart by an inch. Chance (a mere consequence of the act itself) cannot determine desert.

While I do not accept this commonplace of retributivism, my disagreement with Brooks is more fundamental.[15] As I understand criminal desert, it necessarily (that is, as conceptual matter) depends on what goes on elsewhere, both outside the criminal law and within, though not on what *may happen to* go on outside the criminal law or within. Retributivism entails neither methodological individualism nor methodological absolutism. Indeed, as I understand the fairness theory, both are (more or less) ruled out.

For me, the importance of the theory of punishment is *not* showing that punishment is morally permissible. Few doubt that. What is important is providing

practical guidance concerning how much to punish. The centerpiece of my own theory is therefore a procedure for setting penalties. It may be summarized in seven steps:

1. Prepare a list of penalties consisting of those evils (a) which no rational person would risk except for some substantial benefit and (b) which may be inflicted through the procedures of the criminal law.

2. Strike from the list all inhumane penalties.

3. Type the remaining penalties, rank them within each type, and then combine rankings into a scale.

4. List all crimes.

5. Type the crimes, rank them within each type, and then combine rankings into a scale.

6. Connect the greatest penalty with the greatest crime, the least penalty with the least crime, and the rest accordingly.

7. Thereafter: type and grade new (humane) penalties as in step 3 and new crimes as in step 5, and then proceed as above. [16]

Other retributivists have offered similar procedures.[17] I have elsewhere explained why I think mine is better.[18] But what is important now is that I have argued that this procedure can be derived from the basic premises of retributivism (that punishment must be deserved, that the gravity of the offense should determine how much punishment is deserved, and nothing but the gravity of the offense should determine that).[19] That is important because even a quick survey of the seven steps reveals that the procedure is inconsistent with methodological absolutism. The penalty for any particular statutory crime will depend in part on where the end points of the scale are anchored and in part on what other crimes are on the scale. So, for example, the punishment deserved for a particular criminal act will depend in part on whether the maximum statutory penalty is death or only ten years imprisonment, whether the minimum is a small fine or a year in jail, and on how many ranks of crime there are. The act that in Germany would deserve two years imprisonment (or none) may deserve twenty in Texas. The penalties may also

vary with contingent facts about the society. For example, one cannot deserve to have to wear an electronic bracelet as part of house arrest in a society that has yet to invent electronic bracelets or, having them, cannot afford to use them in punishment.

There is a risk of misunderstanding what I just said. I did not say that my version of retributivism says that the punishment meted out for other crimes should affect the sentence given an individual criminal for his crime. I spoke of *statutory* penalties. Once those penalties are set, the crime determines what is (*prima facie*) deserved in punishment. The penalties set for other crimes are irrelevant (just as Brooks claims they should be). I am an "act-absolutist" (though only *prima facie*) without being a "rule-absolutist". Brooks seems entirely unaware of the possibility of such subtleties within a truly retributive theory.

Much the same is true of methodological individualism. I agree with Brooks that only the criminal act should determine the (*prima facie*) punishment. Even society's actual equilibrium is irrelevant to what the criminal deserves for his act (though that consideration, among others, may be relevant for determining whether, all things considered, that punishment or something less should be the actual sentence). My (chief) disagreement with Brooks' individualism concerns statutory penalties. These are, I believe, conceptually linked to society's equilibrium—or, rather, to a part of it, the fair distribution of benefits and burdens. That conceptual linkage is the heart of the fairness theory. The fairness theory explains why it should be—as a part of giving the criminal what he deserves for his crime; and why, therefore, a thoroughly unjust society, such as a concentration camp, cannot have any justified punishments whatever.

Conclusion

This response to Brooks illustrates a conclusion I have come to after several decades working with theories of punishment: *most disagreements concerning the justification of retributivism are disguised disputes about method, not substantive disputes about the theory's merits.* The easiest way to refute a theory is to set the criteria of adequacy so high no theory can satisfy it. Those who, like Brooks, are not simultaneously defending an alternative theory are particularly susceptible to this fallacy.

Not having to keep in mind that a defender of the theory they criticize may respond that their own theory is open to the same objection, they tend to be too free in raising difficulties.

NOTES

[1] T. Brooks, "On Retributivism", *On-Line Philosophy Conference*, Sunday May 21, 2006, p. 12-13. http://experimentalphilosophy.typepad.com/online_philosophy_confere/

[2] Brooks, p. 8.

[3] Brooks, p. 12

[4] Brooks, p. 16.

[5] Brooks, p. 16.

[6] The OED assigns the first use of "retributivism" to a utilitarian, Nigel Walker *(1969)*, a much more recent original than I would have guessed.

[7] See, for example, D. E. Scheid, "Constructing a Theory of Punishment, Desert, and Distribution of Penalties", *Canadian Journal of Law and Jurisprudence* 10 (1997): 441-506; D. Husak, "Why Punish the Deserving?" *Nous* 26 (1992): 447-464; and L. H. Davis, "They Deserve to Suffer", *Analysis* 32 (1971-2): 136-140.

[8] See, especially, H. Morris, "A Paternalistic Theory of Punishment," *American Philosophical Quarterly* 18 (October 1981): 263-271; R. Nozick, *Philosophical Explanations* (Cambridge, MA, 1981), pp. 363-397; J. Hampton, "The Moral Education Theory of Punishment", *Philosophy and Public Affairs* 13 (1984): 208-238; and R. A. Duff, *Trials and Punishments* (Cambridge, 1986)

[9] A. J. Skillen, "How to Say Things with Walls", *Philosophy* 55 (1980): 509-523; I. Primoratz, "Punishment as Language", *Philosophy* 64 (1989): 187-205; and A. von Hirsch, *Censure and Sanctions* (Oxford, 1993). Though J. Feinberg's influential "The Expressive Function of Punishment", *Monist* 49 (1965): 397-423, is usually treated as a seminal statement of the expressive theory—as reprinted in Feinberg's *Doing and Deserving* (Princeton, 1970)—, it is not. It actually proposes to revise the standard (Hart-Flew) definition of punishment, something necessary because of Feinberg's legal positivism. Since retributivism aims at justice, not mere social control, retributive expressivists do not need to fiddle with the definition of punishment to distinguish between (justified and so, expressive) punishment and mere penalties. Retributivists have therefore generally ignored Feinberg's proposal (while citing him as pointing to the expressive function of punishment).

[10] For examples of these unusual approaches to moral retributivism, see P. Montague, *Punishment as Societal Defense* (Lanham, Maryland, 1995), a self-defense approach; and S. Kershnar, *Desert, Retribution, and Torture* (Lanham, 2001), a forfeiture approach.

[11] For an argument of this claim, see M. Davis, "The Relative Independence of Punishment Theory", *Law and Philosophy* 7 (December 1988): 321-350.

[12] Brooks, p. 2.

[13] K. D. Moore, *Pardons: Justice, Mercy, and the Public Interest* (New York, 1989); J.G. Murphy and J. Hampton, *Forgiveness and Mercy* (Cambridge, 1990).

[14] Brooks, pp. 2 and 3.

[15] M. Davis, "Why Attempts Deserve Less Punishment Than Complete Crimes", *Law and Philosophy* 5 (April 1986): 1-32.

[16] M. Davis, How to Make the Punishment Fit the Crime", *Ethics* 93 (July 1983): 726-752.

[17] See, especially, J.D. Mabbott, "Punishment", *Mind* 48 (1939): 152-167 ; and J. Kleinig, *Punishment and Desert* (The Hague, 1973.)

[18] See, especially, M. Davis, *How to Make the Punishment Fit the Crime* (Boulder, 1992).

[19] M. Davis, "How", *Ethics*.