

IS MORAL PHILOSOPHY IRRELEVANT TO THE FORMULATION OF CRIMINAL LIABILITY AND PUNISHMENT RULES?

Paul H. Robinson*

Abstract

Would penal code and sentencing guideline drafters learn anything useful from moral philosophers? Many philosophers won't care if they do or don't, but those who do care may be disappointed because, Robinson argues, what moral philosophers are likely to contribute is both less than and different from what they intend. For a variety of practical reasons, the discipline itself is not well suited to gaining credibility with law and policy makers. More importantly, the law and policy shift toward utilitarianism and away from deontological retributivism seems to take away the major platform from which moral philosophers could have contributed. While it is true that desert enjoys a recent resurgence, that may be based more on its utilitarian advantages than on a newfound deontological interest. And that utilitarian support for desert relies upon an empirical notion of desert -- people's shared intuitions of justice -- not a deontological one. But while one might suppose that a shift to utilitarian desert puts moral philosophers on the sidelines, the truth is that intuitions of justice are of both informal and formal influence in moral philosophy, with the result that the moral philosophy literature is, perhaps inadvertently, the richest source of information in the world today about the subtleties of human intuitions of justice.

Is moral philosophy irrelevant to the formulation of criminal liability and punishment rules? I will argue that much of the moral philosophy literature on criminal law will not be useful and what is useful is probably not what moral philosophers intend to contribute. But let me quickly say that this is not in itself a problem. Many philosophers are happy to be just philosophers, and can be proud of it. (I doubt that my good friend Leo Katz has ever lost sleep worrying about being practically relevant.) But if a philosopher feels a need to make some contribution to real world law and policy making, then my pitch here may be of interest.

Unfortunately, I suspect that many philosophers do want their work to be relevant. It does not take long looking through current philosophy journals to see articles that address real world issues.¹ Philosophers regularly address many of the more timely and debated issues in

* Paul H. Robinson, Colin S. Diver Distinguished Professor of Law, University of Pennsylvania Law School. The author thanks Michael Orchowski for his valuable research assistance.

¹ See, e.g., Leo L. Clarke et al, The Practical Soul of Business Ethics: The Corporate Manager's Dilemma and the Social Teaching of the Catholic Church, 29 Seattle U. L. Rev.139, (continued...)

criminal justice, including everything from complicity and mitigations and defenses to the felony murder rule and the appropriateness of mandatory minimum sentences.² This breadth and the abundance of such publications suggests that many philosophers are interested in justice law and policy and expect their work to have relevance to it.

A Few Practical Barriers to Influence on Law and Policy Making

Law and policy makers tend not to look to moral philosophers for a variety of rather obvious reasons. It's hard to take arguments too seriously if they seem to be offered more as an interesting exercise than as a practical solution, such as attempts to analyze legal doctrine by drawing analogies to chaos theory or other aspects of modern physics.³ Yet this same "interesting game" form of scholarship is well know and regarded in moral philosophy. Philosophers feel entirely comfortable addressing questions in a hypothetical universe distantly

¹(...continued)

203-04 (2005) (promoting the idea that corporations do not exist solely to maximize wealth, and that Catholic social teaching can be used as a lens through which managers may evaluate their social responsibilities); Adam Mossoff, *What Is Property? Putting The Pieces Back Together*, 45 *Ariz. L. Rev.* 371, 443-43 (2003) (arguing that the article's "integrated theory of property . . . explains how eminent domain may have been more varied in its original application, which reveals an implicit conception of regulatory takings that may serve as precedent for today").

² See, e.g., Sharon Dolovich, *Legitimate Punishment in Liberal Democracy*, 7 *Buff. Crim. L. Rev.* 310, 430 (2003) (suggesting, after a lengthy discussion of Rawlsian punishment theory, that mandatory minimum sentences such as California's Three Strikes law may be inappropriate); Sanford H. Kadish, *Respect for Life and Regard for Rights in the Criminal Law*, 64 *Cal. L. Rev.* 871, 897-900 (1976) (identifying the competing doctrines of rights and consequentialism as useful tools for analyzing questions about how human life should be valued by the criminal law and subsequently discussing the relevance of this theory to human medical experimentation, abortion, organ harvesting, and euthanasia); Kyron Huigens, *Homicide in Aretaic Terms*, 6 *Buff. Crim. L. Rev.* 97, 145-46 (2002) (advocating that the aretaic view of homicide necessitates abolition of the felony murder and provocation doctrines); Gerhard Overland, *Self-Defence Among Innocent People*, 2 *J. Moral Philos.* 127, 145-46 (2005) (concluding, after examining the rights of the persons involved, contract theory, incentive arguments and moral fairness, that acts of self-defense are not justified against apparent aggressors unless the actor perceives a show of force that would reasonably suggest that reciprocal force may be necessary).

³ See, e.g., Wai Chee Dimock, *Rethinking Space, Rethinking Rights: Law, Literature, Science*, 10 *Yale J. L. & Human.* 487, 490 (1998) (invoking Einstein's theory of relativity to suggest that rights should be viewed in nonabsolute terms); Rebecca R. French, *Time in the Law*, 72 *U. Colo. L. Rev.* 663, 729-33 (proposing that chaos theory and fractals be used as an alternative means of examining time in the legal field).

removed from reality, where, for example, executing murderers restores the life of the victim.⁴ Other philosophers draw conclusions so contrary to basic intuition that their contentions are difficult to take seriously.⁵ Unfortunately, the popularity of such articles does not build the discipline's credibility in the eyes of law and policy makers.

The absurdity of a premise or conclusion is sometimes obvious, but there are many articles in which the extent of the author's interest in reality is unclear. No doubt philosophers know who among themselves are inclined to deal with the world as it exists, but outsiders may not know this and, therefore, can not easily tell which articles ought to be taken seriously. For many law and policy makers, the only safe course is to avoid the moral philosophy literature altogether.

The discipline's credibility problem is compounded by the fact that it is common to find wide and persistent disagreement among moral philosophers on many if not most significant issues. Indeed, it is likely that on any issue over which law or policy makers themselves disagree, prompting them to look to other disciplines for guidance, philosophers certainly will disagree among themselves.

It is not as if other disciplines don't have disagreements. But, because most have some objective test by which a writer ultimately may be proven right or wrong, there tends over time to be some coalescence around an accepted view. A proposed theory ends up either explaining more of the available data, and is accepted, or does not, and is rejected. But without such a clear test mechanism, moral philosophy lacks a path to coalescence.

Because philosophers will disagree on any significant issue, it is difficult for an outsider to gain something useful, in part because to make an informed judgment of which view ought to be given deference, the outsider must herself know something about moral philosophy. In other words, the outsider must become a bit of an insider. Ultimately, the moral philosophy literature is not terribly accessible; its informed use is commonly costly.

⁴ See, e.g., Gary Colwell, Capital Punishment, Restoration and Moral Rightness, 19 J. Applied Philos. 287, 291 (2002) (contending that capital punishment in a fantasy universe where executing a murderer would restore the life of the victim would be wrong, but that it does not follow that capital punishment is necessarily unethical in the real world)

⁵ See, e.g., Russell L. Christopher, Time and Punishment, 66 Ohio St. L.J. 269, 280 (2005) (arguing that retributivism cannot justify any "divisible" sentence such as imprisonment for a fixed term because one must serve part of a sentence before the full sentence is served, and if the entire sentence is warranted by retributivism, the partial sentence inherently is not; thus, retributivism cannot justify any punishment at all because even something as "indivisible" as, say, a death sentence, can be "divided" into its procedural phases, like spending time on death row, being led to the execution chamber, being bound, and the execution itself, all of which can be considered a part of the "punishment"). David A.J. Richards, Commercial Sex and the Rights of the Person: A Moral Argument for the Decriminalization of Prostitution, 127 U. Pa. L. Rev. 1195, 1286 (1979) (concluding that "the right to engage in commercial sex is one of the rights of the person which the state may not transgress").

Of course, an exposure to the moral philosophy literature may inspire ideas that one might not otherwise have. That is, even if the reasoned conclusions cannot be trusted, the arguments might inspire a reader to think of arguments that she might not otherwise have thought of. But this rather modest contribution – like being a Rorschach stimuli – probably is not what philosophers have in mind if they want their work to have real world influence.

Let me emphasize again that none of this is meant as a criticism but only as an observation. Moral philosophers may have great contributions to make, even if the contributions are not to those who make real world criminal law and policy.

Moral Philosophers in a Utilitarian World

Let us assume for the sake of argument that a moral philosopher could write in a way that is accessible to law and policy makers and is able to gain their attention in a way that distinguishes her from the mass of moral philosophers who are normally ignored. Another major hurdle to having influence is the presently dominant guiding principle for the distribution of criminal liability and punishment. If the guiding principle were deontological retributivism, then such a moral philosopher with credibility would seem to be quite useful to policy makers. But, for good or for ill, the past half century of American criminal law and policy making has been driven by utilitarian analysis, not retributivism.⁶

There is obviously an enormous literature on the subject and a good deal of debate, but I think it is fair to say that most American criminal codes -- based as they are on the Model Penal Code -- are very much concerned with avoiding future crime and, in pursuit of that goal, are willing to adopt rules that many moral philosophers would see as inconsistent with doing justice. Certainly the Model Penal Code itself is an overwhelmingly utilitarian document, in both its provisions and its motivations. Do moral philosophers have a role in a utilitarian world, one in which deontological notions of desert are given limited, if any, value?

Some readers may want to dispute my characterization of modern American criminal law and policy as rejecting deontological notions of desert in favor of utilitarian crime control. After all, there seems to be a recent resurgence of support for distributing criminal liability and punishment according to desert. Witness, for example, the recent ALI drafting committee proposal to set desert as the dominant distributive principle, relegating the traditional crime control mechanisms -- deterrence, rehabilitation, incapacitation of the dangerous -- to a secondary role in which they can be advanced only to the extent that doing so does not conflict with desert.⁷

⁶ Moral philosophers can be utilitarians, of course, but for guidance in utilitarian calculations, law and policy makers are not likely to look to moral philosophy, but rather to other disciplines, such as economics, psychology, sociology, and criminology.

⁷ § 1.02(2). Purposes; Principles of Construction:

(2) The general purposes of the provisions governing sentencing and corrections,
(continued...)

But one may explain this turn of events without assuming that deontological concerns played a role. As I have argued elsewhere, there may be great crime control utility in a desert distribution of criminal liability and punishment, and the shift to a desert distribution may be simply a recognition of the disutility of deviating from a community's shared intuitions about desert.

To quickly summarize those "utility of desert" arguments,⁸ deviating from a community's intuitions of justice inspires resistance and subversion among participants -- juries, judges, prosecutors, and offenders -- where effective criminal justice depends upon acquiescence and cooperation. Relatedly, some of the system's power to control conduct derives from its potential to stigmatize violators -- with some persons this is a more powerful, yet essentially cost-free, control mechanism compared to imprisonment. Yet the system's ability to stigmatize depends upon it having moral credibility with the community; for a violation to trigger stigmatization, the law must have earned a reputation for accurately assessing what violations do and do not deserve moral condemnation. Liability and punishment rules that deviate from a community's shared intuitions of justice undercut this reputation.

Perhaps the greatest utility of desert comes through a more subtle but potentially more influential form. The real power to gain compliance with society's rules of prescribed conduct lies not in the threat of official criminal sanction, but in the influence of the intertwined forces of social and individual moral control. The networks of interpersonal relationships in which people find themselves, the social norms and prohibitions shared among those relationships and transmitted through those social networks, and the internalized representations of those norms and moral precepts control people's conduct.

⁷(...continued)

to be discharged by the many official actors within the sentencing and corrections system, are:

- (a) in decisions affecting the sentencing and correction of individual offenders:
 - (i) to render punishment within a range of severity proportionate to the gravity of offenses, the harms done to crime victims, and the blameworthiness of offenders;
 - (ii) when possible with realistic prospect of success, to serve goals of offender rehabilitation, general deterrence, incapacitation of dangerous offenders, and restoration of crime victims and communities, provided that these goals are pursued within the boundaries of sentence severity permitted in subsection (a)(i) and (ii); and
 - (iii) to render sentences no more severe than necessary to achieve the applicable purposes from subsections (a)(i) and (ii);

...
A.L.I., Model Penal Code: Sentencing, Preliminary Draft No. 3 at 4 (May 28, 2004).

⁸ For a more detailed account, see Paul H. Robinson, & John M. Darley, *The Utility of Desert*, 91 Nw. U.L. Rev. 453 (1997).

The law is not irrelevant to these social and personal forces. Criminal law, in particular, plays a central role in creating and maintaining the social consensus necessary for sustaining moral norms. In fact, in a society as diverse as ours, the criminal law may be the only society-wide mechanism that transcends cultural and ethnic differences. Thus, the criminal law's most important real-world effect may be its ability to assist in the building, shaping, and maintaining of these norms and moral principles. It can contribute to and harness the compliance-producing power of interpersonal relationships and personal morality.

The criminal law also can have effect in gaining compliance with its commands through another mechanism. If it earns a reputation as a reliable statement of what the community perceives as condemnable, people are more likely to defer to its commands as morally authoritative and as appropriate to follow in those borderline cases in which the propriety of certain conduct is unsettled or ambiguous in the mind of the actor. The importance of this role should not be underestimated; in a society with the complex interdependencies characteristic of ours, an apparently harmless action can have destructive consequences. When the action is criminalized by the legal system, one would want the citizen to "respect the law" in such an instance even though he or she does not immediately intuit why that action is banned. Such deference will be facilitated if citizens are disposed to believe that the law is an accurate guide to appropriate prudential and moral behavior.

The extent of the criminal law's effectiveness in all these respects -- in avoiding resistance and subversion of an unjust system, in bringing the power of stigmatization to bear, in facilitating, communicating, and maintaining societal consensus on what is and is not condemnable, and in gaining compliance in borderline cases through deference to its moral authority -- is to a great extent dependent on the degree to which the criminal law has gained moral credibility in the minds of the citizens governed by it. Thus, the criminal law's moral credibility is essential to effective crime control, and is enhanced if the distribution of criminal liability is perceived as "doing justice," that is, if it assigns liability and punishment in ways that the community perceives as consistent with their shared intuitions of justice.⁹ Conversely, the system's moral credibility, and therefore its crime control effectiveness, is undermined by a distribution of liability that deviates from community perceptions of just desert.

The important point here is to see that the moral philosophy conception of a desert distribution is not only unnecessary for these utilitarian crime control benefits but indeed would be ineffective in gaining them. The beneficial consequences of a desert distribution, listed above, flow not from following a deontological desert distribution but only from following a desert distribution of an empirical sort -- one that tracks the community's shared intuitions of justice. It is the community's *perception* that justice is being done that pays dividends, not the system's actual success as measured by deontological notions of desert.

This empirical conception of desert is different from the deontological conception in more than just its rationale. The difference in rationales can generate important practical differences in results. An obvious example is found in the significance of resulting harm. Moral

⁹ For documentation and discussion of shared intuitions of justice among lay persons, see Paul H. Robinson & Robert Kurzban, *Intuitions of Justice* (forthcoming, draft available from author).

philosophers disagree about its significance -- whether, all other things being equal, resulting harm increases blameworthiness and deserved punishment -- and each side of the debate has plausible arguments to make.¹⁰ In contrast, all available data suggests a nearly universal and deeply held view among the community that resulting harm does matter, that it increases the offender's deserved punishment.¹¹ This practical difference is only one of a host of issues on which a moral philosopher's conclusion may vary from the empirical data on shared intuitions of justice.¹²

But one might argue that the recent ascendance of desert as a distributive principle can be explained either as a shift toward deontological desert or toward utilitarian desert. But the evidence, while not entirely clear, does not seem to support the claim. First, as noted above, the results of the two conceptions of desert can be different, and it is the empirical notion that typically prevails, not the deontological version. So, for example, the law reform trend has been toward unanimity in giving significance to resulting harm. Second, today's law and policy makers (perhaps other than those philosophically trained) tend to give people's-sense-of-justice explanations for desert-based legislation, empirical claims, not deontological arguments.¹³

¹⁰ Compare, e.g., Leo Katz, *Why the Successful Assassin Is More Wicked than the Unsuccessful One*, 88 Calif. L. Rev. 791, 806 (2000) (arguing by hypothetical that moral logic suggests that harm should be considered when assessing blameworthiness) and Michael S. Moore, *The Independent Moral Significance of Wrongdoing*, 5 J. Contemp. Legal Issues 237, 267-71 (1994) (positing that our own experiences, in that we resent successful wrongdoers more than we do those who unsuccessfully attempt harm, we feel more guilty about our own completed misdeeds than we do about attempts, and we are dissatisfied with reasonable moral choices that produce undesirable consequences suggests that "results matter" in the moral arena) with Richard Parker, *Blame, Punishment, and the Role of Result*, 21 Am. Phil. Q. 269, 273 (1984) (advocating that resulting harm should not be relevant to punishment determinations, as "[f]ortune may make us healthy, wealthy, or wise, but it ought not determine whether we go to prison.").

¹¹ See Paul H. Robinson & John M. Darley, *Justice, Liability, and Blame: Community Views and the Criminal Law*, 14-28, 181-96 (1995). Mansfield, for example, concludes that "[t]he notion that there should be a difference in punishment [between unsuccessful attempts and completed crimes] is deeply rooted in popular conscience, and to ignore it is to risk jury nullification." John H. Mansfield et al, *Causation in the Law – A Comment*, 17 Vand. L. Rev. 487, 494-95 (1964).

¹² For community views on a wide variety of criminal law issues, see Robinson & Darley, *supra* note 11, which also compares community views to existing legal rules and discusses the points of disagreement.

¹³ See, e.g. A.L.I. Model Penal Code: Sentencing, *supra* note 7, at 9-10 (suggesting that because desert can be difficult to quantify at times, sentencing commissions should solicit a diverse range of community perspectives and that doing so gives the commissions a "unique credibility.").

The Moral Philosopher's Role in Assessing Utilitarian Desert

Does this mean, ironically, that the recent ascendance of desert as a dominant distributive principle leaves moral philosophers as irrelevant as they were before its rise? From what I have said above, one might think this to be the case. It is the empirical notion of shared intuitions of justice that matters, not the moral philosophy conception of deontological desert. But, in truth, I think the ascendance of desert, even an empirical notion of desert, is a turn that makes the moral philosophy literature quite important.

First, for empiricists to test shared intuitions of justice and to model the same, they must have some conceptualization with which to start; they must have some hypothesis of what distinctions are important that they can then test. With an initial model, they can test its pieces and see what adjustments, if any, need be made. Is there reason to think that the moral philosophy literature can help them in this work? I think yes. Indeed, it offers the only existing thoughtful and detailed examination of what constitutes principles of desert from which such a starting model can be constructed.

Second, it is likely that much of the moral philosophy literature has been influenced by, and therefore is a goldmine for learning about, people's intuitions of justice. Being human, philosophers are likely to share in human intuitions of justice. (Those intuitions can be lost through intellectualization. They often are lost by many in the law and economics school, but some may cite the frequency of goofy moral philosophy articles as evidence that the same loss is possible among philosophers.) But even more than philosophers sharing in human intuitions of justice, there is a substantial tradition in moral philosophy of giving importance to intuitions of justice. A common form of analysis and debate is to argue by hypothetical.¹⁴ Implicit in this form is the belief that people can agree on how some hypotheticals should come out and that this agreement means that consensus conclusion is reliable as a philosophical matter. If questioned, the debaters may quickly note that agreement is not conclusive, but as a practical matter agreement among those present does pass as conclusive proof of the correctness of that disposition. Thus, one would expect that over time philosophical positions developed through the use of such argument by analogy would have a bias toward conclusions consistent with shared intuitions of justice. Principles that might have logical support, nonetheless might gain

¹⁴ See, e.g., Leo Katz, *Incommensurable Choices and the Problem of Moral Ignorance*, 146 U. Pa. L. Rev 1465, 1482183 (1998) (using a hypothetical, derived from application of necessity defense to situations where the actor has culpably created the justifying situation, to argue that at times persons can be blamed for making the wrong decision in a state of “unavoidable moral ignorance”). I have argued elsewhere that this kind of argument by analogy is dangerous because it commonly violates basic methodology rules for accurately testing a person's true intuitions of justice. See Paul H. Robinson, *Some Doubts About Argument by Hypothetical*, 88 Cal. L. Rev. 813, 823 (2000) (“[In some cases,] [t]he results we get . . . are probably not intuitive judgments of blameworthiness but more likely intellectualized answers generated by applying the professor's resident collection of theoretical positions -- [for example,] whether resulting harm ought to be judged significant”).

little currency or be discarded because philosophers as a group think they produce unjust results – a practical veto by shared intuitions.

Often a good deal of energy is spent constructing a theory that seeks to give a logical explanation for a shared intuition. The more the shared intuition varies from the position of simple logic, the greater the investment of intellectual energy to justify it. Witness the debates on the significance of resulting harm in assessing deserved punishment, noted previously.¹⁵ Every empirical test confirms the dramatic commitment among lay persons to the notion that resulting harm increases blameworthiness and deserved punishment. Yet opponents can offer what they see as a strong argument that it is absurd to think that a person's blameworthiness can be affected by circumstances outside of his control. We may properly give greater punishment to one who tries to cause a harm than to one who only risks causing a harm, they explain, but the luck of whether the harm actually comes about is not within the actor's control and ought not play a role. An actor can control only his conduct and state of mind and only these ought to affect his blameworthiness. Obviously there are a full set of counterarguments to make but they are beside the point. The point is that the matter is long debated and has produced a large literature, in part, I suspect, because intuition and logic seem so opposed.¹⁶

But it is not just that shared intuitions of justice are part of the background influence on philosophers as they do their work. The reliance upon shared intuitions has in many quarters been given formal authoritative status. Consider the mechanism for determining principles of justice embodied in the "original position" and the "veil of ignorance" advanced by Rawls and others.¹⁷ It is a thought experiment that essentially formalizes the reliance upon one's intuitions of justice, and tries to control how that intuition testing is to be done, in a way that will exclude factors thought improper, such as self-interest.

The philosophical test mirrors quite nicely what social psychologists would do when testing intuitions of justice upon which a utilitarian-desert criminal code could be based. They construct scenarios that ask the test subject to judge blameworthiness in a case in which the subject has nothing at stake, and they manipulate the scenarios to avoid the presence of factors, such as the race or class of the actors, that they fear might distort the subject's blameworthiness judgment. A psychology research methodologist would feel quite pleased with Rawls' construction of test conditions through the "original position" and the "veil of ignorance."

Conclusion

With such a tradition of giving weight to intuitions of justice, both formally and informally, it should be no surprise that researchers will find moral philosophy a rich source of distinctions that are important to shared intuitions of justice, as well as an invaluable source of a

¹⁵ See text accompanying note 10, *supra*.

¹⁶ See articles listed in note 10, *supra*. I find some of the argument in support of the significance resulting harm quite clever and persuasive, but I also know that I thought resulting harm ought to count even before I heard these clever arguments.

¹⁷ See, e.g., John Rawls, *A Theory of Justice* 118 (1971).

suitable conceptual framework to use as a starting point for modeling human judgements of moral blameworthiness.

But the contribution that philosophers make in this way may not necessarily be the one that they intend. It is not that their logical arguments are persuasive, or even relevant. Rather, their work is being used more cynically: its usefulness comes not from its adherence to logic and sound reasoning but rather from its perhaps inadvertent influence by philosophers' personal intuitions.

Whether the contribution is as intended or not, moral philosophy does have a contribution to make. No one else has thought so long and hard about the components of moral blameworthiness. If desert utilitarians ignore the moral philosophy literature in their construction of the empirical notion of desert, they are ignoring the most thoughtful body of literature on the subject that exists today.