The Scope Problem in Punishment

Diana Mertz Hsieh (diana@dianahsieh.com)
Philosophy of Law (Phil 5260, Boonin)
11 May 2006 (Spring 2005)

The proper moral justification of the institution of legal punishment, whereby the state punishes persons found guilty of criminal offenses, is not self-evident. Philosophers of law usually attempt to justify it by appeal to one or more of the major benefits supposedly thereby gained: deterrence of future crimes, incapacitation of the criminal, rehabilitation of the criminal, and/or retribution for the crime. The resulting theories of punishment often suffer from serious defects, including an inability to properly distinguish between offenders and non-offenders. In such cases, instead of justifying the punishment of all and only criminal offenders, the theory permits the punishment of some innocent persons and/or forbids the punishment of some guilty persons.

This paper will argue that theories of punishment based upon deterrence, incapacitation, or rehabilitation inevitably violate the proper scope of punishment in various ordinary cases by failing to justify the punishment of all and only criminal offenders. That's because those justifications for punishment appeal to some future good rather than to any facts about the offense itself. Some of the scope problems are only magnified by appealing to all three of these future benefits of punishment, as utilitarian theories usually do. Ultimately, these problems are so grave that utilitarian theories of punishment cannot be justly regarded as genuine theories of punishment at all.

The Problem of Punishment

In its broadest sense, punishment refers to "a penalty imposed for wrongdoing," whether cutting off the hand of a man guilty of stealing or grounding a teenager for breaking his curfew.[1] Legal punishment, in contrast, is restricted to the penalties imposed by the state upon persons found guilty of criminal offenses. The "problem of punishment" raises the question of whether legal punishment can be justified—and if so, by what means. As David Boonin explains in the opening passage of his draft book *The Problem of Punishment*:

> Legal punishment involves treating those who break the law in ways that it would be wrong to treat those who do not. Even if we assume that those who break the law are responsible for their actions and that the laws that they break are just and reasonable ones, this practice raises a problem. How can the mere fact that a person is responsible for having broken a just and reasonable law render it permissible to treat him in ways that it would otherwise be impermissible to treat him? How can the line between those who break such laws and those who do not be morally relevant in the way that the practice of punishment takes it to be? This is the problem of punishment.[2]

So the problem of punishment concerns the morality of state punishment of criminal offenders per se—not whether given class of actions should be crimes (e.g. prostitution), nor the limits of personal responsibility for crimes (e.g. criminal insanity), nor the proper standards for determining guilt (e.g. reasonable doubt), nor just forms of punishment (e.g. death penalty). Such questions are temporarily set aside in the problem of punishment, so as to focus on the core

[1] http://dictionary.reference.com/search?q=punishment
[2] David Boonin, "The Problem of Punishment," (2006), 1.

cases of some punishment by the state of persons known to be willful offenders of just laws. In essence, the problem of punishment asks: What, if anything, morally justifies the punishment of criminal offenders by the state? The answer determines the basic purpose and character of a criminal justice system.

Contemporary philosophers of law attempt to justify legal punishment by appealing to some benefit of punishment, whether deterrence of future crimes, incapacitation of the criminal, rehabilitation of the criminal, retribution for the crime, or some combination thereof. Utilitarian theories generally invoke the first three (i.e. deterrence, incapacitation, and rehabilitation) as the positive benefits to be gained by state punishment of criminal offenders.[3] Although other costs must be considered in the moral calculus, those future benefits render the intentional harms of some punishments morally permissible, if not obligatory. In contrast, retributive theories claim that punishment is morally grounded in the fact that a criminal offender deserves to be punished for his crimes.[4] Although other benefits of punishment are often recognized as genuine, only desert morally justifies punishment. Thus the basic distinction between utilitarian and retributive theories of punishment concerns the moral purpose served by legal punishment: social benefits in the future or just compensation for past offenses.

The Scope Problem in Punishment

One standard requirement of any theory of punishment is that the theory must properly distinguish between law-breakers and law-abiders, morally justifying the punishment of the former while forbidding punishment of the latter. A theory unable to meet this "scope requirement" would permit (if not demand) the punishment of some innocent persons and/or forbid the punishment of some guilty persons. Notably, such scope problems are ontological, not epistemological; the state is presumed to know whether the person punished or not is actually guilty or not. As we shall see, utilitarian theories of punishment are plagued by worries about the scope of punishment. Their advocates often attempt to sidestep such worries altogether by one of two argumentative strategies—and those must be excluded at the outset.

First, some utilitarians appeal to the definition of punishment, claiming that no state act qualifies as legal punishment unless the person convicted of some criminal offense is actually guilty.[5] So the idea of punishment of the innocent is "a logical impossibility" unable to serve as an objection to any theory of punishment.[6] Even if that's true, even if the problem of punishing the innocent should be instead described in terms like "false punishing" or "pseudo-punishing," the moral problem remains the same as ever.[7] The proposed theory of punishment still would justify the infliction of the very same intentional harms as punishment by the very same state's criminal justice system upon persons innocent of any crime.[8] It would permit (if not demand) treating law-abiding citizens as criminal offenders—and vice versa. The grave moral objections to the theory generating those results do not evaporate by the mere "verbal maneuver" of

[3] David M. Adams, "Introduction to Punishment," in *Philosophical Problems in the Law*, ed. David M. Adams (New York: Wadsworth, 1996), 414., S. I. Benn and R. S. Peters, "The Utilitarian Case for Deterrence," in *Philosophical Problems in the Law*, ed. David M. Adams (New York: Wadsworth, 1996), 416-7., Richard Brandt, "The Utilitarian Theory of Criminal Punishment," in *Readings in Philosophy of Law*, ed. John Arthur and William H. Shaw (Englewood Cliffs, NJ: Prentice-Hall, 1995), 190.
[4] Adams, "Introduction to Punishment," 414.
[5] Benn and Peters, "The Utilitarian Case for Deterrence," 418.
[6] Ibid.
[7] Lecture Notes, Philosophy of Law, January 27, 2005
[8] Lecture Notes, Philosophy of Law, January 27, 2005

claiming that an innocent man sentenced to 20 years in prison by a criminal court is, technically speaking, not punished.[9] So this first strategy fails.

Second, some utilitarians attempt to "bite the bullet" by defending some punishment of the innocent and some non-punishment of the guilty as well-justified by the overarching ethical theory of utilitarianism. On this approach, "our ordinary convictions about punishment for crime ought to be thoroughly re-examined in important respects" so that we might understand the moral propriety of punishing law-abiders but not law-breakers when that yields the most desirable outcomes.[10] Apart from utilitarianism's well-known weaknesses as a moral theory, the basic problem is that abandoning the requirement of punishment for all and only those guilty of criminal offenses means abandoning the institution of punishment itself. Under such a system, the state would intentionally harm various people for the sake of reducing crime, without regard for the guilt or innocence of those harmed. Such harms would be punishment in name only, since they would be only accidentally connected to any criminal offenses.[11] Similarly, when a teacher gives grades students based upon her personal feelings, she abandons the practice of grading their work, even though she might tend to prefer the diligent students. So the second strategy fails too.

Although these two strategies for dismissing scope problems in punishment may seem wholly opposite, they both capitalize on a failure to properly distinguish the descriptive and normative senses of "punishment." The descriptive sense simply refers to the full range of intentional harms to persons actually sanctioned by a given theory of punishment, whereas the normative sense of punishment refers to the intentional harms that a state rightly imposes upon all and only those persons guilty of criminal offenses. In the first strategy, the utilitarian attempts to ignore the untoward implications of his own theory by using only the normative sense of punishment. He dismisses any deviations by his theory from the norms of punishment as something other than punishment—and so not his concern. In the second strategy, the utilitarian attempts to embrace whatever his theory demands by appealing to the descriptive sense alone. So whatever his theory demands is punishment, even if that violates the basic norms thereof. In fact, a proper theory of punishment must align the descriptive and normative meanings of "punishment," such that all the punishments (descriptive) sanctioned by the moral theory are actually proper punishments (normative). That's why the scope of punishment, i.e. whether a theory permits the punishment of all and only offenders, is a critical test for any theory of punishment.

So let us now consider whether any of the three utilitarian-flavored theories of punishment—i.e. those based upon deterrence, incapacitation, and rehabilitation—can avoid scope problems by justifying the punishment of all and only criminal offenders. Since deterrence is the major variation thereof, we shall examine it in far greater detail than the other two options. Finally, we shall consider whether the fully utilitarian hybrid of all three theories fares any better than its parts.

Punishment as Deterrence

In deterrence theories of punishment, the state punishment of criminal offenders is morally justified by that institution's capacity to discourage criminal offenses: punishments for past crimes show potential offenders the harms to be inflicted upon those convicted of future

[9] Brandt, "The Utilitarian Theory of Criminal Punishment," 193.
[10] Ibid.
[11] Lecture Notes, Philosophy of Law, February 1, 2005

crimes. Punishment "serves to deter potential offenders by inflicting suffering on actual ones."[12] The basic goal of deterrence-based punishment is not to restrain or reform offenders, but rather "to give potential offenders prudential reason to obey the law" in the form of fear of suffering the harms of punishment.[13]

Utilitarian theories of punishment generally regard deterrence as the major positive value of punishment.[14] Moreover, deterrence theories are typically some form of utilitarianism, if not always so. In that framework, legal punishment must maximize social welfare, meaning that the benefits gained by deterring crime must be weighed against the various costs of punishment. Such costs almost inevitably include the expense of punishment, the fear aroused in potential offenders of punishment, and even the suffering of the punished criminal himself. They may also include the harms of any unjust punishment—such as punishing offenders excessively, punishing innocent persons, and failing to punish guilty persons—or even fears thereof in the general population. Absent consideration of such costs, deterrence theory would sanction absurdly harsh punishments as the most effective deterrents of crime, such that even a first-time teenage shoplifter of a candy bar could be tortured to death. The utilitarian framework of deterrence theory precludes such grossly disproportionate punishments, at least in most ordinary cases.

Deterrence theories generally adopt the "rule" rather than the "act" form of utilitarianism.[15] So punishments are justified in advance for various types of crimes, as opposed to on-the-spot for this or that particular crime. For instance, Joe would be punished with 15 years hard labor for his murder of Mary based upon the general deterrence value of that kind of punishment for that kind of crime, not based upon any expectation that his particular punishment will deter four to six murders. As Benn and Peters write in "The Utilitarian Case for Deterrence": "It is not a question of what conditions a particular act of punishment must satisfy, but of the conditions that a *rule* must satisfy if punishment is to be properly attached to it."[16] However, deterrence theorists (like all rule utilitarians) must somehow explain why people working in the justice system ought not occasionally and quietly defect from those general rules when doing so would maximize social welfare (or deterrence). After all, although our current legal system is heavily governed by rules, it's not clear that it ought to remain so. And even within our present system, police, prosecutors, judges, and juries often exercise substantial discretion. So it's far from obvious that a rigidly rule-bound system of the sort often advocated by rule utilitarians as necessary would, in fact, be better than all the alternatives.

So can a deterrence theory of punishment satisfy the discrimination requirement that all and only criminal offenders be punished by the state? In short, the answer is no. Absent an ad hoc rule that all and only criminal offenders may be punished, deterrence theories will permit the state to punish the innocent and/or not permit the state to punish the guilty in various ordinary cases. The basic reason is that deterrence of future crime requires only the incitement of fear in people about the harms imposed by the state for such crimes, yet (1) such fears may be aroused by harming those known to be innocent or wrongly thought guilty and (2) such fears may not be aroused by harming those known to be guilty or wrongly thought innocent. Deterrence theory

---

[12] Benn and Peters, "The Utilitarian Case for Deterrence," 417.
[13] Antony Duff, *Legal Punishment* (Spring 2004) (The Stanford Encyclopedia of Philosophy, Spring [cited); available from http://plato.stanford.edu/entries/legal-punishment/.
[14] Benn and Peters, "The Utilitarian Case for Deterrence," 417.
[15] Ibid., 419., Brandt, "The Utilitarian Theory of Criminal Punishment," 193.
[16] Benn and Peters, "The Utilitarian Case for Deterrence," 419.

will permit the punishment of the innocent in the first cases and forbid the punishment of the guilty in the second cases.

In the first set of cases, deterrence theory must sanction the deliberate punishment of innocent persons when doing so would deter crime without incurring costs greater than the benefits of that deterrence. Some cases would require deception by the state: the innocent person punished would be widely thought guilty. For example, when the person guilty of a crime seems unlikely to be identified or convicted, evidence could be suppressed and/or manufactured to frame an innocent man. So long as most people believe the frame-up, that unjust punishment will deter potential offenders more than the alternative of punishing no one for the crime. Even the real culprit, presumably still on the loose, may be deterred from committing more crimes in the hope that laying low will allow him to remain undetected and unpunished. The basic problem is that deterrence depends upon subjective beliefs: fear of punishment for crimes will be aroused so long as people think the punished person guilty, whether he actually is or not. Consequently, the punishment of an innocent man widely believed to be guilty may sometimes deter crime more effectively than the punishment of a guilty man widely believed to be innocent. In fact, if the justice system is generally but wrongly regarded as incompetently punishing the innocent while allowing the guilty to go free, then potential offenders likely will underestimate the risk of punishment for their own crimes. In that case, the state could better deter crime by sometimes if not routinely punishing those thought guilty, whether actually guilty or not.

The standard response by the deterrence theorist to such objections is that these cases fail to account for all the costs of punishing the innocent, particularly the "disastrous effects on public confidence" if such unjust punishments were ever brought to light.[17] However, if revealed, such punishments probably could be effectively portrayed as honest errors by the courts—or just the doings of a corrupt government official or two. Whistleblowers from within the justice system would be unlikely, since those people would be committed to deterrence as the moral purpose of punishment. Moreover, if most ordinary people understood deterrence to be the basic aim of punishment, then revealing such deceptions ought not cause any great public fury. Since the punishments of the known innocents would no longer serve any deterrent effect, those people could be released from punishment (if possible) without undue fuss. All in all, the deterrence theorist cannot offer a decisive case to support his claim that the secret punishment of the innocent in place of the guilty always or even usually incurs more costs than benefits.

These cases of secret unjust punishment suggest a far more insidious problem for the deterrence theorist, namely that a criminal justice system primarily committed to deterring future crimes probably could not commit to the various procedures and safeguards designed to prevent the conviction and punishment of innocent persons. No grand conspiracies that might deal a "shattering blow to public confidence and security" would be required, merely a gradual loosening of standards for physical evidence, testimony, interrogation, and the like.[18] That would permit the punishment of innocent persons as a matter of course, not based upon any intentional deception but rather simple negligence. In essence, a deterrence-based criminal justice system might simply grow lazy about the guilt or innocence of the persons punished by it.

In addition to such quiet punishments of the innocent *instead of* the guilty, deterrence theory also plausibly sanctions the open punishment of the innocent *in addition to* the guilty. To take a fairly standard example, some (potential or actual) criminals would be substantially less inclined to commit crimes if conviction resulted not just in their own punishment, but also in

[17] Brandt, "The Utilitarian Theory of Criminal Punishment," 193.
[18] Ibid.

punishment of their loved ones.[19] So a man convicted of murder could be obliged to watch the execution of his wife or girlfriend. If he is convicted of battery, she might just be beaten. The prospect of such "auxiliary punishment" would likely deter many potential offenders more than just the prospect of their own punishment. Also, the potential innocent victims of such punishment then would be strongly motivated to keep the offender on the straight and narrow, perhaps even reporting any suspicions to the police if that would protect them from any auxiliary punishment.

Of course, the deterrence theorist will quickly observe that any system of auxiliary punishment would incur substantial costs, particularly in the form of the fears and sufferings of all the people associated with criminals. However, such costs likely would be straightforwardly offset by the additional deterrence of future crime made possible by auxiliary punishment. So imagine a criminal justice system in which auxiliary punishments would consist of basically the same crimes as those committed by the offender, such as rape for rape, murder for murder, assault for assault. In that case, the costs associated with any given act of auxiliary punishment would be basically identical to the costs of permitting a single crime of the same type as the criminal offense—or of failing to deter it. In both the auxiliary punishment and the undeterred crime, the innocent victims would be harmed in fairly similar ways. For example, imagine that a state-appointed punisher batters Mary as auxiliary punishment for her hot-tempered husband's battery of a bystander in a bar fight—and that deters two other hot-tempered men from committing the same crime. That's a net gain in social welfare, since now just one innocent person is battered rather than two. Unlike those two potential victims attacked suddenly, Mary will likely have some time to dread the forthcoming suffering of her punishment. Yet that might not be psychologically worse for her in the long run than a sudden attack from nowhere. Moreover, knowing the limits of the attack, e.g. that the battery will not escalate into rape or murder, might also be of substantial comfort to Mary. So by deterrence theory, Mary could plausibly be punished for the crimes of another person—without need for the complications of secrecy. In general, so long as a system of auxiliary punishment would deter more crimes than it committed (so to speak), then a deterrence theory would be obliged to allow it as morally permissible, if not obligatory.

Similarly, deterrence theory would likely sanction the collective punishment of a whole group of people when some unknown member of that group is known to be guilty.[20] So if money is stolen from the bank account of a business to which only four employees had access, then the state might simply punish all four people for embezzlement. Unlike the punishment of the innocent *instead of* the guilty, such punishment of the innocent *along with* the guilty would deter just as much crime as the punishment of the guilty party alone. It might even deter more crime if the potential criminal cares enough for the other members of the group likely to be punished with and for him. Such collective punishment would undoubtedly deter more crime than refraining from punishment entirely on the grounds that the particular guilty person cannot be identified. However, as with other forms of punishing the innocent, a system of collective punishment would come at a price, particularly in the form of mistrust or resentment between persons. Still, the utilitarian cannot plausibly argue that such costs clearly outweigh the benefits of so deterring crime.

---

[19] C. L. Ten, "Fantastic Counterexamples and Utilitarian Theory," in *Philosophy of Law*, ed. Joel Feinberg and Hyman Gross (New York: Wadsworth, 1995), 630.

[20] Benn and Peters, "The Utilitarian Case for Deterrence," 420n2.

We've so far examined three kinds of cases in which deterrence theory would plausibly sanction the punishment of the innocent: punishment of framed innocents, punishment of loved ones, and punishment of groups. In the second set of cases, deterrence theory plausibly fails to justify some punishment of the guilty on the grounds of its failure to adequately deter crime. Notably, deterrence theory does not seem to suffer from so many grave problems on this score as with the punishment of the innocent.

In *Introduction to the Principles of Morals and Legislation*, early utilitarian Jeremy Bentham claims that the inherent evils of punishment cannot be justified in various kinds of cases, including when the prospect of punishment "could produce no effect on [the offender], with respect to the preventing him from engaging in any act of the sort in question."[21] Bentham then describes one such form of "inefficacious" punishment:

> Where, though the penal clause might exercise a full and prevailing influence, were it to act alone, yet by the *predominant* influence of some opposite cause upon the will, [punishment] must necessarily be ineffectual; because the evil which [the offender] sets himself about to undergo, in the case of his *not* engaging in the act, is so great, that the evil denounced by the penal clause, in case of his engaging in it, cannot appear greater. [22]

Bentham is explicitly concerned with cases in which fear of some physical danger, whether due to natural causes or human coercion, can only be avoided by some crime—and the penalties for it do not outweigh the evils of the physical danger itself.[23] However, these same considerations could well apply to people whose basic goals and desires in life can only be satisfied by criminal activity. This category of criminals might include child molesters with twisted desires to sexually dominate another, mobsters revolted by the prospect of honest work, and spiteful people hell-bent upon avenging the wrongs that supposedly ruined their lives. Such criminals may be so sociopathic (even though not criminally insane) that they will plan, scheme, and ultimately risk any punishment whatsoever rather than abandon their criminal activities. They would attempt to avoid punishment, but only by concealing their crimes, never by ceasing to commit them. Notably, such criminals need not be punished in order to deter others from committing such crimes, since people without the relevantly-twisted psychology would be in no danger of descending into rape, thuggery, and vengeance. The crime would be so distasteful in and of itself to the vast majority of people that it would serve as its own deterrent. Since the prospect of punishment would not deter those crimes, punishment could not be justified by deterrence alone. (A more broadly utilitarian theory incorporating incapacitation and/or rehabilitation would be required.) To be fair, the deterrence theorist might plausibly argue that the prospect of punishment for such crimes slows down the offender's rate of criminal activity by forcing him to take measures to avoid detection. However, that's more plausible in cases of often-repeated crimes (e.g. sexual molestation) than in cases of a single, long-planned crime (e.g. vengeance murder).

In addition, since effective deterrence depends solely upon the widespread belief in the punishment of the guilty, deterrence theory may not always require the actual punishment of the guilty but rather only sometimes the public impression thereof.[24] If the criminal is unlikely to repeat his crime, he might be "punished" with a fake public torture or banishment to some pleasant Caribbean island. A famous scientist sentenced to ten years in prison for killing his

---

[21] Jeremy Bentham, *An Introduction to the Principles of Morals and Legislation* (Oxford: Clarendon Press, 1907), XIII.9.
[22] Ibid., XIII.11.
[23] Ibid.
[24] Lecture Notes, Philosophy of Law, February 3, 2005

unfaithful wife might be instead sent to work in a secret government lab. A reformed murderer on death row might willingly forsake the long and costly appeals process to live out the rest of his life in moderate comfort in a distant land after faking a public execution. In general, such false punishment would be acceptable by the standards of deterrence theory so long as (1) the criminal himself would not repeat his crime and (2) the criminal would be silent about his non-punishment if ever returned to society. While the evils of not punishing the guilty may not loom quite so terribly for deterrence theory as the evils of punishing the innocent, they also cannot be dismissed as insignificant.

So as we've already seen, a theory of punishment must squarely confront scope problems: the punishment of the innocent and/or non-punishment of the guilty cannot be rightly dismissed as irrelevant or embraced as proper. That leaves the deterrence theorist with just one argumentative strategy: asserting that the substantial benefits apparently gained in various cases of punishing the innocent or not punishing the guilty are always outweighed by countervailing costs. Too often that's a highly implausible claim on empirical grounds alone, particularly in cases of punishment of the innocent. As C. L. Ten observes in "Fantastic Counterexamples and the Utilitarian Theory," in disputes about the proper utilitarian moral calculations about punishment "no one can claim with confidence that, on the balance of probabilities, there are not actual cases in which punishing the innocent will produce the best consequences."[25] Such debates about the moral calculus are inevitable because deterrence theories (and other utilitarian theories of punishment) cannot ever rule out the punishment of the innocent or non-punishment of the guilty as a matter of principle. The punishment of the innocent is always a live option because punishment of only those guilty of crimes is not necessarily the most effective deterrent, just one method thereof. And the non-punishment of the guilty is always a live option because the punishment of those guilty of crimes is not necessarily a deterrent at all.

The fundamental problem with deterrence theory is that the state can fairly easily arouse fear of the consequences of crime based upon (1) false beliefs about the identity of the offender and/or the harms inflicted upon him and (2) the prospect of harms to innocent people near and dear to the offender. And so punishment based upon deterrence will not be inflicted upon all and only offenders—except by accident. The deterrence theorist might attempt to categorically block all scope problems by imposing the rule that all and only criminal offenders are to be punished. Yet that would be ad hoc and arbitrary, basically equivalent to declaring that no brown-eyed persons will be punished. Similarly, the deterrence theorist cannot coherently appeal to any moral principles (such as showing "respect for persons" or acting as "a responsible person") that might ever conflict with deterring crime in particular or maximizing social welfare in general.[26] Ultimately, the best that the advocate of deterrence theory can say is that the moral calculus doesn't work out in favor of punishing the innocent or not punishing the guilty—for now.[27] Yet even that's implausible, as we've seen.

Punishment as Incapacitation

In incapacitative theories of punishment, the state punishment of criminal offenders is morally justified by that institution's capacity to prevent future crimes via physical restraints such as execution, incarceration, and castration. Incapacitative punishments work upon the body only; they do not attempt to change the offender's state of mind by instilling fear or inspiring

[25] Ten, "Fantastic Counterexamples and Utilitarian Theory," 630.
[26] Benn and Peters, "The Utilitarian Case for Deterrence," 418-9.
[27] Ten, "Fantastic Counterexamples and Utilitarian Theory," 630.

reform.[28] Although most punishments involve some form of incapacitation, purely incapacitative punishments might not be punishments at all, but rather more like quarantines of the sick for the danger they pose to others.[29] If the goal of incapacitation is merely to prevent the criminal from committing more crimes, then the harms inflicted upon that criminal by incapacitation would be incidental, not intentional, as required for punishment. Still, perhaps the unmitigated frustration of the criminal's desires could be an intentional harm of incapacitation. In any case, because incapacitation is regarded as a major benefit of punishment, particularly by utilitarians, the question of the scope of punishment on an incapacitative theory should be at least briefly discussed.[30]

Due to its focus on preventing future crimes, incapacitative punishment would both permit punishment of the innocent and forbid punishment of the guilty. Most obviously, if punishment of criminal offenders is morally justified by a rational expectation of future crimes, then punishment of purely potential offenders could also be justified by a rational expectation of future crimes. So a man who routinely drinks to excess, verbally rages at his wife, and punches walls might be preemptively incarcerated for spousal abuse, even though he's never physically attacked or even threatened his wife. That would be the punishment of the (criminally but not morally) innocent. Conversely, criminal offenders could not be punished for their past crimes absent some reasonable expectation of their committing future crimes. So a woman who quietly murders her rotten husband then leads an exemplary life for the next twenty years, never remarrying nor even dating, could not be punished for that murder because she's probably no more likely to commit a crime now than any other random woman.

More generally, a system of incapacitative punishment might cast a wide net so as to catch *all* those who would otherwise commit crimes in the future. (That might require incapacitating everyone.) Or it might cast a narrow net so as to catch *only* those who would otherwise commit crimes in the future. (That might require incapacitating no one.) However, no solid middle ground of punishing all and only future criminal offenders can be found, since individuals have the power to choose whether to engage in criminal activity or not—up to the very last moment before the crime. The probability of future offenses cannot be derived from even perfect knowledge of a person's past actions and present character. So no system of incapacitation can cast an accurate net that catches *all and only* those who would otherwise commit crimes in the future. That's likely one of the reasons why incapacitation is only rarely upheld as its own theory of punishment, but rather generally serves as part of a broadly utilitarian theory.

Punishment as Rehabilitation

In rehabilitative theories of punishment, the basic purpose of punishment is the moral improvement of the offender such that crime is no longer an active option in his deliberations about his future actions. Punishment is justified by the possibility of the future benefit of moral reform, meaning a change in "offenders' dispositions [so] that they will in future willingly obey the law."[31] The reformed criminal would not merely realize "crime does not pay," but also "show… remorse for his past misdeeds, and resolve… not to repeat them, not through fear of

[28] Duff, *Legal Punishment* ([cited).
[29] Ferdinand Schoeman, "On Incapacitating the Dangerous," in *Readings in Philosophy of Law*, ed. John Arthur and William H. Shaw (Englewood Cliffs, NJ: Prentice-Hall, 1995), 217.
[30] Benn and Peters, "The Utilitarian Case for Deterrence," 416.
[31] Duff, *Legal Punishment* ([cited).

further punishment, but simply because they were wrong."[32] For rehabilitation to be a form of punishment, the sufferings involved in punishment must be necessary for the reformation of moral character.[33]

So would a system of rehabilitative punishment err by punishing the innocent or failing to punish the guilty? Yes. As with incapacitation, a person might be preemptively rehabilitated while still innocent of any crime. So the verbally-abusive, wall-punching drunken husband could be incarcerated for spousal abuse—and also rehabilitated for it. If that does not seem terribly unjust, consider that such people are clearly capable of mending their ways without the coercive interference of the state. If the state steps in to punish them, however, they lose the opportunity to do that—and are thereby intentionally harmed in ways to which they might never consent.

In addition to sanctioning the punishment of such innocents, rehabilitative punishment also precludes the punishment of some people guilty of criminal offenses. For example, if Lance is caught burgling houses but reforms his character before trial—perhaps by apologizing and compensating his victims, studying the importance of property rights, and training himself for honest work as a paralegal—he could not be punished for his past crimes. According to a rehabilitative theory, he has already achieved the basic purpose that such punishment would serve. Far worse, a criminal known to be thoroughly incorrigible—perhaps due to explicit statements by him of his wholehearted commitment to his criminal activities—could not be punished, since rehabilitation would be impossible.

The basic reason for such scope problems in rehabilitation theories is that punishment is based upon a person's present and future character rather than his past actions. If that character needs reform and might be reformed, then punishment is justified. If that character does not need reform or cannot be reformed, then punishment is not justified. Of course, the fact that a person committed a crime is a good indicator that he is of bad character, but the correlation is not perfect in either direction. As with incapacitative theories of punishment, such obvious scope problems are probably a major reason why rehabilitation is often just an element of a broader theory of punishment.

Hybrid Utilitarian Punishment

Thus far, we've examined the three major future benefits of punishment—deterrence, incapacitation, and rehabilitation—to determine whether any might adequately serve as the moral purpose of punishment. All have been found wanting simply based upon the inability of the resulting theory of properly distinguish between offenders and non-offenders by punishing all the former and none of the latter. As already indicated, utilitarian theories of punishment generally appeal to all three of these benefits to morally justify punishment, usually with the heaviest emphasis upon deterrence. However, this hybrid approach does not solve the problems of scope, but merely compounds them.

In a utilitarian theory of punishment, each of the three future benefits of punishment (i.e. deterrence, incapacitation, and rehabilitation) would serve as an independent source of moral justification for punishment. So if punishing Jake would both deter other criminals and incapacitate him without significant overriding costs, then that punishment would be justified, even if the criminal was not reformed in the slightest. (The same method would apply to the justification of rules.) That method of independent sufficient justifications would sometimes

[32] Benn and Peters, "The Utilitarian Case for Deterrence," 417.
[33] Ibid., 416.

help overcome the problem of not punishing the guilty, insofar as the three possible benefits provide additional possible avenues of justification for punishment. So if the punishment of Joe for the murder of Mary cannot be justified by rehabilitation because he's already reformed his character, he might still be punished on the grounds that his punishment might deter other potential murderers. However, those additional avenues of justification greatly compound the problem of punishing the innocent, since innocent persons may be justified on the basis of deterrence *or* incapacitation *or* rehabilitation. So all discussed forms of punishing the innocent could be justified by a utilitarian theory.

Alternatively (and contrary to utilitarianism), imagine that each of the three future benefits of punishment (i.e. deterrence, incapacitation, and rehabilitation) were individually necessary and jointly sufficient conditions of morally justified punishment. In that case, the punishment of innocent persons would be far more difficult, since it would require deterrence *and* incapacitation *and* rehabilitation. Such a theory would forbid the auxiliary punishment of the loved ones of criminals, for example, since that would neither incapacitate nor rehabilitate the criminal offender. However, this approach to punishment would greatly magnify the failures to punish the guilty of the three individual theories, since a criminal offender would have to be effectively deterred, incapacitated, and rehabilitated to justify punishment. That last demand would be particularly difficult to meet, since moral reform is substantially dependent upon a person's voluntary choices. So criminals would routinely go unpunished.

Ultimately, utilitarian theories of punishment are not genuine theories of punishment at all. In part, that's because they cannot meet the basic normative requirement for any theory of punishment that the state punish all and only criminal offenders. More fundamentally, those scope problems are the result of utilitarianism's requirement that punishment be justified by reference to some future benefits to be gained by it, not by any consideration of the wrongness of the prior offense. As a result, utilitarian theories of punishment do not intentionally inflict harms upon *criminals* for their *offenses* but rather intentionally inflict harms upon *anyone* to *prevent future crime*. As such, they are properly considered theories of crime reduction, not theories of punishment. In contrast, retributive theories of punishment are focused upon the past criminal offense—and so the sufferings imposed upon the criminal by the state would at least count as punishment for those past offenses. Of course, whether retributive theories of punishments suffer from other kinds of defects—and some versions surely do—must be a subject for another time.

Works Cited

Adams, David M. "Introduction to Punishment." In *Philosophical Problems in the Law*, edited by David M. Adams, 414-16. New York: Wadsworth, 1996.

Benn, S. I., and R. S. Peters. "The Utilitarian Case for Deterrence." In *Philosophical Problems in the Law*, edited by David M. Adams, 416-20. New York: Wadsworth, 1996.

Bentham, Jeremy. *An Introduction to the Principles of Morals and Legislation*. Oxford: Clarendon Press, 1907.

Boonin, David. "The Problem of Punishment." 2006.

Brandt, Richard. "The Utilitarian Theory of Criminal Punishment." In *Readings in Philosophy of Law*, edited by John Arthur and William H. Shaw, 189-95. Englewood Cliffs, NJ: Prentice-Hall, 1995.

Duff, Antony. *Legal Punishment* (Spring 2004) The Stanford Encyclopedia of Philosophy, Spring [cited. Available from http://plato.stanford.edu/entries/legal-punishment/.

Schoeman, Ferdinand. “On Incapacitating the Dangerous.” In *Readings in Philosophy of Law*, edited by John Arthur and William H. Shaw, 214-23. Englewood Cliffs, NJ: Prentice-Hall, 1995.

Ten, C. L. “Fantastic Counterexamples and Utilitarian Theory.” In *Philosophy of Law*, edited by Joel Feinberg and Hyman Gross, 618-31. New York: Wadsworth, 1995.