University of Chicago Legal Forum

Volume 2005 | Issue 1 Article 11

Expressing Doubts about Expressivism

Heidi M. Hurd
Heidi.Hurd@chicagounbound.edu

Expressing Doubts About Expressivism

Heidi M. Hurd[†]

There are five theories that are common coins in exchanges about how punishment can best be justified. And there are some profound problems that confront each of these theories that have motivated criminal law theorists to seek new justifications for the state's ability to deprive citizens of their liberty, and sometimes their lives. The contender in vogue today is expressivism, which finds punishment a promising means of contradicting the "social meanings" of crimes and declares the communicative function to be the highest and best goal of the justice system. My questions in this Article are these: Can expressivism carve out a space for itself that is genuinely separate from the other theories that have long competed for dominance in criminal law theory? And if it can lay claim to philosophical autonomy, does expressivism provide a conceptually coherent and morally attractive alternative to its classic competitors?

As a means of getting some purchase on these questions, I shall first outline the five traditional goals that have been set for the criminal law and provide a thumbnail sketch of the problems that pose the greatest obstacles to theorists who seek to defend them. This will enable us both to measure the theoretical space that is left for a sixth alternative and to assess the promise of expressivism as a contender.

I shall suggest in Part II that the reliance placed by expressivism on the notion of crimes and punishments having "social meanings" renders it conceptually troubled, if not incoherent, and the moral difficulties that an expressivist theory of punishment invites are reminiscent of those raised in response to several of the more traditional theories upon which it is supposed to be an improvement. When one takes seriously the significance of the conceptual and moral problems encountered by attempts to defend an expressive justification of punishment, one cannot claim that it poses a formidable challenge to the traditional theo-

[†] Dean and David C. Baum Professor of Law and Philosophy, University of Illinois.

ries that have long vied for dominance in criminal law jurisprudence.

I. FIVE TRADITIONAL THEORIES OF PUNISHMENT

A. Corrective Justice Theory

Let us begin with the aims of the corrective justice theorist (sometimes described as the restorative justice theorist[1]), whose concern is with employing criminal punishment to compensate or restore victims who have been wronged by offenders. According to the corrective justice theorist, punishment of an offender is justified if and only if it vindicates the offender's victim.[2] Corrective justice is the animating theme behind the new victims' rights movement that has asserted itself throughout Western nations in recent years, and it is the source of many proposed procedural changes in the criminal justice system, such as the admission of victim impact statements and victim testimony at sentencing.[3]

The most pressing question for a corrective justice theory is this: Does it have any internal checks on the amount of punishment imposed on an offender, or will it be vulnerable to charges that it will license gross over- and under-punishment? In two words, the problem is summed up as follows: Lorena Bobbit! Suppose that what it takes to vindicate a victim of domestic abuse is castration, torture, or death. Is there anything internal to a corrective justice theory that checks the conclusion that if this is what it takes, then this is what ought to be administered?

The corrective justice theorist might answer by insisting that whether a victim is vindicated or not depends not upon whether she is *subjectively* restored to psychic repose, but upon

---

[1] See, for example, Heather Strang and John Braithwaite, eds, *Restorative Justice: Philosophy to Practice* (Ashgate 2000).

[2] See, for example, Kathleen Daly, *Revisiting the Relationship between Retributive and Restorative Justice*, in Strang and Braithwaite, *Restorative Justice* at 36 (cited in note 1) (stating that, in the context of contrasting restorative justice to "traditional justice practices," the emphasis in restorative justice "is on repairing the harm between the offender and the victim").

[3] See Barbara Hudson, *Victims and Offenders*, in Andrew von Hirsch, et al, eds, *Restorative Justice and Criminal Justice: Competing or Reconcilable Paradigms?* 177, 179–80 (Hart 2003) ("All versions of restorative justice have at their centre the opportunity provided for the victim to recount what the offence meant to her."); Allison Morris and Warren Young, *Reforming Criminal Justice: The Potential of Restorative Justice*, in Strang and Braithwaite, eds, *Restorative Justice* at 15 (cited in note 1) (stating that a strong presence for the victim in criminal justice processes is central to restorative justice).

whether her offender is forced to disgorge the *objective* benefits that he unjustly reaped at her expense. On this argument, while a victim's blood-thirst might know no bounds, and so provides no limits to an offender's punishment, an offender's unjust enrichment both is susceptible to objective measure and happily coincides with what he justly deserves to lose. Hence, if punishment deprives a wrongdoer of his ill-gotten gains, then it vindicates the victim while ensuring that the wrongdoer gets no more than he deserves.

There are four related and significant problems with this version of corrective justice that any corrective justice theorist worth her salt is going to have to address. First, in what sense is a corrective justice theory, construed as the objectivist would construe it, different from a retributive justice theory, which holds that the punishment of an offender is justified if and only if the offender deserves the punishment?[4] If victims are (objectively) vindicated if and only if their offenders receive their just deserts (as measured by whether their unjust enrichment was eradicated)—regardless of whether victims feel vindicated or are in any appreciable way psychologically restored as a result of the punishment—then it would seem that corrective justice theory is not an autonomous theory of punishment, for it cannot justify anything other than what is otherwise justified by retributivism.

Second, how, and by what objective measure, is an offender unjustly enriched by a crime? It is easy to characterize a theft as a crime of unjust enrichment: one who steals a million dollar Ming vase has unjustly enriched himself by a million dollars, and his victim is vindicated when the vase or its value has been restored to him. But how are we to characterize and measure the unjust enrichment of the kidnapper, the traitor, the rapist, and the vandal?

Third, in what sense could the imprisonment of such offenders go any distance toward returning their victims to their ex ante positions? One understands how dollars could go some distance towards achieving compensation, since dollars have some ability—albeit imperfect—to buy back what might have been stolen from the victim through the crime (her health, psychic repose, lost earnings, and so on). But, of course, this is the task of tort law. If corrective justice theory is going to constitute a theory of criminal punishment that does not, by its own logic, spell the collapse of criminal law into tort law, then it is going to have

[4] See Part I-E for a discussion of the retributive theory of punishment.

to explain how an offender's years in prison disgorge to his victim something that goes some distance toward making her whole again, and that does so in a manner that cannot better be accomplished through tort damages.

Finally, this task of connecting punishment to just compensation might be an intelligible one if the corrective justice theorist were to invoke a *subjective* measure of the punishment required to achieve the vindication of the victim. Then, the question would be analogous to the one that is, in principle, asked in tort law in order to fix the appropriate level of damages: namely, at what point would the plaintiff have been indifferent between the harm done to her and a cash payment? It is the plaintiff's subjective indifference curve that provides the most promising measure of when and how a given damage award makes her whole, such that she is no worse off by virtue of having been harmed and compensated than she would have been had the harm (and ensuing compensation) never occurred. But if there comes a point at which a person would be indifferent between keeping her hand and receiving dollars for its loss, it is hard to imagine that there is any meaning in the suggestion that persons (at least those devoid of masochistic tendencies) would ever be indifferent between retaining bodily security and seeing someone punished for its loss.

Of course, even if we can solve some of the problems that beset a corrective justice theory by returning to a subjective measure of the punishment necessary to achieve a victim's vindication, we are returned to the problem that imposing punishment proportionate to the subjective needs of victims invites the prospect of gross over-punishment, for victims may not rest easy until their assailants are made to suffer from quite Draconian measures.

And just as a subjective measure permits over-punishment, so it also permits what most would think of as intolerable under-punishment. Suppose that a victim denies any need for vindication. Citing the virtue of turning the other cheek, she implores the court not to punish her own brutal rapist. In the face of her manifest lack of need for vindication (indeed, in the face of her manifest need to recover her dignity by being merciful to one who showed her no mercy), ought the court to let her assailant go free? It would seem that there is simply no getting around the haunting problem that corrective justice theories, by their attention to what it would take to permit the victim the kind of psychic repose she enjoyed prior to the offense, appear to license punishments disproportionate to desert.

B. Rehabilitation Theory

On a rehabilitation theory, punishment is justified if and only if it accomplishes a metamorphosis in the personality of the offender so as to make him a socially functional participant in the moral community.[5] The concern of the rehabilitationist is the moral transformation of the offender, and the roots of his theory concerning the justifiable goals of the criminal justice system lie deep in a perfectionist theory of legislation. On such a theory, the power of the state is properly directed toward the goal of perfecting persons morally. Perfectionists consider it legitimate for the state to enact legislation that will make us more virtuous and less vicious by nurturing in us charitable, kind, and courageous dispositions, and suppressing selfish, cruel, and bigoted dispositions. In the context of criminal law, this general political philosophy inclines theorists to the view that suffering can be justified if, but only if, it is motivated by the promise of improving the moral character of offenders.

It is crucial to be clear at the start that anyone who seeks to rehabilitate offenders because it will deter them from further wrongdoing and so make the rest of us safer, or because it will return them to the pool of productive, contributing citizens and so increase the wealth and well-being of all, is *not* a true rehabilitationist. Such a theorist is a utilitarian who conceives of the rehabilitation of offenders as an instrumentally effective means of maximizing social utility overall, but who is not principally interested in the moral welfare of the offender. In contrast, a true rehabilitationist is someone who takes the rehabilitation of the offender to be an intrinsic good, and who is prepared to sacrifice social utility in order to achieve that end.

We can thus imagine two systems of punishment that equally deter crime. One does so by Draconian means, making punishment so gruesome that would-be offenders are frightened out of acting on their wicked motivations. The other does so by rehabilitating offenders, purging them of their desires to pursue

---

[5] See, for example, Herbert Morris, *A Paternalistic Theory of Punishment*, in Jeffrie G. Murphy, ed, *Punishment and Rehabilitation* 154–68 (Wadsworth 3d ed 1995) (arguing that the moral good of the wrongdoer ought to count heavily in the moral legitimacy of punishment); Karl Menninger, *The Crime of Punishment* 17–18 (Viking 1969) (characterizing crimes as analogous to diseases and criminals as in need of cures, not "justice"); B.F. Skinner, *Beyond Freedom and Dignity* 81 (Knopf 1971) (arguing that the task of the criminal justice system "is to make life less punishing and in doing so to release for more reinforcing activities the time and energy consumed in the avoidance of punishment").

criminal ends and substituting motivations of a virtuous sort.[6] Imagine that the second system is three times more expensive than the first. The utilitarian will opt for the first, for her ends are thereby accomplished through minimally costly means, thus maximizing social utility summed over all citizens. The rehabilitationist will opt for the second, for only it accomplishes the ends that he believes are justifiably pursued by the criminal justice system.

The prospect that true moral rehabilitation would be far more expensive than other means of deterrence is hardly unrealistic, given how difficult it is for any of us to change our stripes. But costliness, while of substantial legitimate concern to the utilitarian (and to anyone who takes it to be a side-constraint on the systematic pursuit of justice) is not the only problem that threatens the rehabilitation theory.

Consider the moral of the following story. In 1958, John Lynch was convicted of indecent exposure for engaging in masturbation while at a drive-in hamburger joint.[7] That offense was treated as a simple misdemeanor, eligible for no more than brief jail time or a small fine.[8] In 1967, he was again convicted of the same offense under similar circumstances. This time he was eligible for a life sentence under California Penal Code § 314.[9] California had pioneered indeterminate sentencing schemes motivated by an "enlightened" desire to rehabilitate offenders, rather than to punish them, and had legislated an indeterminate sentence of one year to life for second-offense indecent exposure.[10] The state's theory was that offenders should spend as much or as little time in prison as was necessary to alter their unfortunate dispositions. As the California Supreme Court in *In re Lynch* recognized, this scheme, and the rehabilitative theory that moti-

---

[6] This thought experiment is drawn from Michael S. Moore, *Law and Psychiatry; Rethinking the Relationship* 234–35 (Cambridge 1984), reprinted in Leo Katz, Michael S. Moore, and Stephen Morse, eds, *Foundations of Criminal Law* 60, 61 (Oxford 1999).

[7] *In re Lynch*, 503 P2d 921, 939–40 (Cal 1972).

[8] Cal Penal Code § 314 (West 1970). Mr. Lynch received two years of probation for his first offense. *Lynch*, 503 P2d at 922–23.

[9] Cal Penal Code § 314 (West 1970), repealed in relevant part by the Uniform Determinate Sentencing Act (1976), codified primarily at Cal Penal Code § 1170 and in other various sections (West 2004 & Supp 2005). At the time of Mr. Lynch's second offense, § 314 read, "Upon the second and each subsequent conviction [for indecent exposure], every person so convicted is guilty of a felony, and is punishable by imprisonment in state prison for not less than one year." Cal Penal Code § 314 (West 1970). See also *People v Caddick*, 160 Cal App 3d 46, 51–52 (1984) (describing the history of indeterminate sentencing in California).

[10] Cal Penal Code § 314; Cal Penal Code § 1168.

vated it, would readily permit someone who had twice exposed himself to serve many times more years in prison than someone who committed manslaughter, robbery, mayhem, and many different kinds of violent assault[11]—not because indecent exposure is a grave wrong, but because its motivations may be difficult to purge[12] (as § 314 of the California Penal Code clearly anticipated). If it would indeed have taken twenty-five years to alter Mr. Lynch's proclivities, then the rehabilitationist can have no complaint about incarcerating him for twenty-five years in response to two episodes of public exposure!

Lynch teaches the lesson that rehabilitationism has no internal check on the *amount* of punishment that can be imposed on offenders, save its efficaciousness in altering character. Rehabilitationism thus permits incarceration that may far exceed (or fall far short of) what an offender deserves. Now consider a tale that illustrates that the theory has no internal check on the nature of treatment that can be delivered, save efficacy of character modification.

In the 1970s, staff at the Vacaville facility in California employed the drug anectine as a means of achieving aversion therapy.[13] Anectine paralyzes the involuntary motor muscles that control the lungs, and so generates an acute and quite terrifying sense of suffocation. After receiving anectine injections in tandem with lectures on the wickedness of their ways, offenders found that they would become breathless, very anxious, and sometimes phobic in any circumstance that invited the prospect of their recidivism. Their rule infractions dropped by 30 percent,[14] proving anectine a powerful means of altering the motivations of would-be recidivists. Inasmuch as character is simply a function of motivation, those who would embrace a rehabilitationist theory of punishment must seemingly regard such drug-therapy with favor. Their compelled enthusiasm, however, reveals that not only does rehabilitationism lack checks on the amount of punishment that can be imposed; it also lacks checks on the kind of punishment that can be imposed.

---

[11] Id at 935–38.

[12] Id at 925.

[13] See Roy G. Spece, Jr., Note, *Conditioning and Other Techniques Used to "Treat?" "Rehabilitate?" "Demolish?" Prisoners and Mental Patients*, 45 S Cal L Rev 616, 633–38 (1972) (describing the program at Vacaville and the effects of anectine).

[14] Id at 636–37.

C. Utilitarian Theory

The utilitarian's guiding concern is to maximize social utility summed across all members of society. Her theory thus takes punishment to be justified if and only if the pain or disutility imposed upon the offender is outweighed by a net gain of utility to others.[15] The utility that may be achieved by punishing offenders can take many forms, and it is commonplace for people to advance several reasons for punishment without clearly appreciating that they all presuppose a commitment to a utilitarian theory. An offender's incarceration by itself can prevent further wrongdoing during the duration of the incarceration, and it is commonly hoped and claimed that incarceration is sufficiently unpleasant to deter recidivism after an offender's release. The punishment of an offender may also effectively deter other would-be offenders by giving proof that crime does not pay. Punishment may serve to educate those who are motivated to be law-abiding but who also need clear guidance concerning the law's demands so as to deter them from crime less by threat than by education. The imposition of pain on those who have caused pain may prove cathartic for the many who are outraged at the criminal act, and so may yield increased utility. Those who would otherwise resort to vigilante justice may be deterred from taking the law into their own hands if the law is quick to avenge their losses. And finally, punishment may reaffirm the social contract, providing assurances to those who sacrifice liberty by abiding by the law that they are not being played for fools by free-riders who would have the gains of cooperation without the reciprocal costs.

Given how terribly common claims of these kinds are amongst theorists and laypersons alike, it is important for criminal law theorists to acknowledge the problems that beset the theory of punishment upon which they rely. By its very terms, a utilitarian theory is committed, in principle, to permitting the punishment of an innocent person and the acquittal of a guilty assailant. For if the visible punishment of a person known by a judge to be innocent will deter those who would be eager to offend, educate those who are morally ignorant, reinforce the social

---

[15] See Jeremy Bentham, *An Introduction to the Principles of Morals and Legislation* 170–71 (Prometheus 1988) ("Upon the principle of utility, if [punishment] ought at all to be admitted, it ought only to be admitted in as far as it promises to exclude some greater evil."); Richard A. Posner, *An Economic Theory of the Criminal Law*, 85 Colum L Rev 1193, 1205–14 (1985) (finding that criminal liability is imposed for acts that violate the rules of consensual transfers).

contract between those whose law-abidingness depends upon it, reduce vigilantism by those tempted by it, and achieve other social gains that collectively outweigh the costs of undeserved punishment to the person falsely accused, then the utilitarian theory will demand such punishment. Only doubt about predicted consequences slows the utilitarian. In a case in which the consequences are clear, the right thing to do is whatever will yield a net gain in social utility. Hence, where a judge can predict that social utility will be maximized by punishing an innocent person, he should punish the innocent and take cost-efficient means of preventing knowledge of his deed.

Similarly, of course, when social gains may be achieved by allowing a guilty person to go free, the utilitarian takes no offense at recommending that result. Thus, if greater racial harmony was achieved by the acquittal of O.J. Simpson, and that benefit outweighed the social disutility that accrued from the decision, then the utilitarian would applaud the outcome of that famous criminal case, even if she were convinced that Simpson was in fact guilty of the murders of Nicole Brown Simpson and Ron Goldman. Once again, only uncertainty concerning consequences gives a utilitarian pause: nothing inherent in the theory in principle precludes gross under- and over-punishment.

Indeed, the theory commits one to thinking that the optimal state of affairs would be achieved if the guilty were indeed apprehended, tried, and condemned very publicly, but then secretly given new identities in happy circumstances or treated to an enviable existence in a remote tropical location. Were we successfully to pretend to punish the guilty (while in fact treating them to a life that they would prefer), we would achieve all the gains that the utilitarian seeks from punishment without any of the disutility that punishment imposes on the offender. Inasmuch as the disutility of the offender is of moral relevance to the utilitarian, her theory, by its own terms, favors a reduction (and, if possible, an elimination) of punishment when its gains can be achieved through other, less costly means—most plausibly through the ruse of punishment.

D. Mixed (Retributive/Utilitarian) Theory

The fourth philosophical perspective on punishment represents an attempt to repair the problem raised for utilitarianism. It is the perspective of the so-called "mixed theorist," who takes an offender's just deserts to cap the amount of punishment legitimately imposed on the offender, but who insists that punish-

ment is justified only if it *also* accomplishes a net gain in social utility. On this theory, then, utility is the accelerator and desert is the brake. Persons may not be punished if they are undeserving, and they may not be punished more than they deserve, but they ought not to be punished if no social gain can be achieved through their pain.

One hears expressions of this theory quite commonly, when, for example, people ask what good it will do to punish someone who is old, ill, or repentant. Very elderly Nazi war criminals, who have only recently been discovered to have lived innocuous lives for decades under assumed identities, often garner surprising sympathy from people who simply see no social utility in punishing them all these years after their youthful atrocities.

While the mixed theory escapes the prospect of theoretically vindicating the punishment of the innocent, it surely fails to correct for the other two interrelated problems that, in my view, defeat the theoretical defensibility of utilitarianism. Suppose that a brutal rapist finds God and is born again, so as to believably pose no threat of returning to his criminal ways. Suppose also that the mere pretense of his punishment will suffice to accomplish the other goals that might be socially useful: the deterrence of other would-be rapists, the reinforcement of the social contract, the moral education of the general public, and so on. Should he nevertheless be punished?

The answer of the mixed theorist must be "no." If no good will come of his punishment, or if the good that would come of it can be accomplished through pretending to punish him rather than actually punishing him, then the right thing to do is to let him go free (and indeed, to invest cost-efficient resources in helping him to do so, so as to maximize utility summed across all members of the community, including the rapist). So the problem is this: a mixed theory of punishment does not demand that offenders get their just deserts. It simply demands that they get no more than their just deserts.

E. Retributive Theory

The retributivist takes punishment to be justified if and only if it accords an offender his just deserts.[16] On this theory, desert is both necessary and sufficient for punishment. Punishment

---

[16] Immanuel Kant, *The Metaphysical Elements of Justice* 100–02 (Bobbs-Merrill 1965) (John Ladd, trans); Moore, *Law and Psychiatry* at 236 (cited in note 6). See also Herbert Morris, *On Guilt and Innocence* 33–36 (California 1976).

should never be imposed on one who is undeserving, or in an amount that is undeserved; but when punishment is deserved, no greater gain need be sought than to administer it in proportion to the culpable wrongdoing of the offender.

This theory clearly answers the concerns about under- and over-punishment raised by the previous theories, for it makes desert the touchstone of justifiable sanctions. Those who doubt that it is an *intrinsic* good to accord persons their just deserts may continue to be skeptical about the justifiability of devoting resources toward the punishment of the guilty absent larger social gains, but such a complaint simply returns critics to the task of defending the utilitarian or mixed theory presupposed by their skepticism.

Retributivism is not without its puzzles, however, and it seems to me that the greatest one is the question of whether it can defend, in any principled way, claims about the proportionality of particular punishments to the particular harms culpably caused by particular offenders.

The principle of Lex Talionis is a powerful answer to the challenge that punishments cannot be matched to offenses: it straightforwardly demands an eye for an eye, a tooth for a tooth, and a life for a life. Surely to get back what you give is the most intuitive notion of getting your just deserts.

But no retributivist would embrace the claim that the state ought generally to be in the business of perpetrating the horrors on offenders that match in kind the horrors that they perpetrated on their innocent victims. No one in today's academy believes that the state ought to satisfy the demands of retributivism by torturing the torturer, raping the rapist, or flashing the flasher. And what would it even mean to be treasonous to the traitor?

Yet once the retributivist leaves a principle that exactly matches punishment to offense, what allows the retributivist to say that seven months, or seven years, or seventeen years, is the just reward of one who commits tax fraud? What principle works to match differential losses of liberty with crimes of differential severity?

The traditional answer of the retributivist is to reject the claim implicit in this criticism that punishments must be absolutely proportionate to their punishments. Instead, the answer goes, punishments are justified if they are *comparatively* propor-

tionate to their crimes.[17] Accordingly, if crimes are rank-ordered from most serious to least serious—from murder to double-parking—and if punishments are then assigned to each crime in descending order of severity, then it is immaterial whether tax fraud gets seven years or seventeen years, so long as it gets the same punishment as identically-serious crimes, less punishment than the next most serious crime, and greater punishment than the next less serious crime. That seventeen years in prison may not match our Platonic conception of what tax fraud merits is not a legitimate complaint if embezzlement receives twenty years and mail fraud fourteen years, and tax fraud is proportionately more serious than the latter and less serious than the former; and so on with regard to all other crimes, from the least serious to the most serious.

Perhaps by repairing to comparative proportionality the retributivist can allay fears of moral arbitrariness. Certainly, to the extent that just punishment is largely a function of equal treatment, a system of comparative proportionality will ensure that like cases are treated alike, and different cases are treated proportionately differently.

But I find the retributivist's often blithe assumption that comparative proportionality will rescue him from arbitrariness to be optimistic at best, and hypocritical at worst. Imagine a society that articulates and enforces very few criminal prohibitions. It makes it criminal to murder, rape, kidnap, commit theft, engage in vandalism, or defame another's good name (considering it as serious to "steal" another's reputation as to steal his television). It condemns murderers to death, gives a life sentence for rape, and gives fifty years for kidnapping, forty-five years for theft, forty years for defamation, and thirty-five years for vandalism. Could a gossip legitimately complain when imprisoned for forty years for falsely calling her neighbor dishonest? It would seem not, at least if the test of just punishment is whether the crime is punished proportionately to other crimes. For while it would seem, in absolute terms, extraordinary to be locked away for the bulk of one's productive life for a petty slur, one cannot complain that the punishment is disproportionate to other punishments. But inasmuch as the moral force of the retributivist's theory lies

---

[17] See, for example, *State v Benn*, 845 P2d 289, 324 (Wash 1993) (holding that Benn's sentence was comparatively proportionate to other sentences for similar crimes). See also John Kleinig, *Punishment and Desert* 115–16 (Martinus Nijhoff 1973) (describing the argument that punishments are just if crimes are scaled from most to least deserving of punishment and then matched to a ranking of the harshness of punishments).

in its concern for matching punishments to just deserts, it seems a fatal flaw to then declare that judgments of comparative proportionality are close enough for government work!

II. EXPRESSIVISM: A VIABLE ALTERNATIVE?

The chronic inability of punishment theorists to solve the problems inherent in the five traditional perspectives on punishment that I have outlined has motivated a number of contemporary criminal law theorists to seek new philosophical foundations for the institution of punishment. Today's most fashionable contender is that of the self-described expressivist, who maintains that punishment is justified if and only if it properly denounces or contradicts the social meaning of an offender's actions.[18] On this theory, we are to view criminal activity as expressing a message or having what is called "social meaning," and we are to view the role of punishment as contradicting that message. Both crime and punishment are thus thought to be communicative exercises. As Dan Kahan has written,

> On expressive grounds serious crimes strike us as such—that is, as *crimes* and as *serious*—not just because they impair another's interests, but because they convey that the wrongdoer doesn't respect the true value of things. To express condemnation, then, society must respond with a form of punishment that unequivocally evinces the community's repudiation of the wrongdoer's valuations.[19]

One who is attracted to this new expressivism has the task of answering both conceptual and moral questions, and it is to these questions that I want to devote the remainder of this article. Conceptually, the expressivist must be able to make sense of what it means to say that criminal deeds and their punishments have "social meanings," such that punishments can be employed strategically to contradict the insults communicated by offenses (Part A below). Morally, the expressivist must be able to motivate the claim that communicating messages that contradict those of criminals is intrinsically good (not just instrumentally

---

[18] See, for example, Dan M. Kahan, *The Anatomy of Disgust in Criminal Law*, 96 Mich L Rev 1621 (1998) (arguing that persons' "disgust sensibility" explains many otherwise puzzling facets of criminal law—such as the fact that imprisonment is the preferred alternative for punishment in America—because they are centered on the question of what acts deserve to be the objects of social expressions of disgust).

[19] Id at 1641 (emphasis in original).

good) (Part B below), and is so good that it sometimes justifies punishing persons more than they deserve (Part C below), and less than they deserve (Part D below).

A. Making Conceptual Sense of the "Social Meaning" of Crimes and Punishments

The first question that an expressivist must answer is this: How are we to extract the message embodied in a given criminal action so as to know how to craft punishment in a way that will contradict it? One might begin by hypothesizing, plausibly enough, that crimes can be thought to have what H.P. Grice termed "natural meaning": they are symptoms of natural phenomena to which they are causally related.[20] Just as we say that a sneeze "means" that one has a cold, a clap of thunder "means" that a storm is brewing, and a loud exclamation "means" that one is surprised, so we might think that a crime "means" whatever it is to which it is causally connected.[21] On this conception, seeking the meaning of a crime is like seeking the meaning of red spots on someone's skin: one does not inquire into anyone's intentions; rather one seeks to determine phenomena of which these effects are symptomatic. Thus one might say that "the spots mean measles,"[22] and "his crime means he was jealous, or his wife left him, or unemployment breeds desperation." In neither case, however, has one characterized either his spots or his deeds as communicative acts, despite their being what Grice termed natural "signs."[23]

It is not worth denying that crimes can be signs in the Gricean sense. Indeed, they evidence a great many important background causes and so might take their "natural meaning(s)" from any and all those contributors—economic, social, psychological, physiological, and so forth. But expressivists could hardly equate the social meanings of crimes with their natural mean-

---

[20] H.P. Grice, *Meaning*, in Jay F. Rosenberg and Charles Travis, eds, *Readings in the Philosophy of Language* 436, 437 (Prentice-Hall 1971).

[21] This evidential construal is articulated by one perspicuous critic of expressivism, Matt Adler. See Matthew D. Adler, *Expression and Appearance: A Comment on Hellman*, 60 Md L Rev 688, 708 (2001); Matthew D. Adler, *Expressive Theories of Law: A Skeptical Overview*, 148 U Pa L Rev 1363, 1384–85 (2000). For a general discussion of nonlinguistic meaning, see Heidi M. Hurd, *Sovereignty in Silence*, 99 Yale L J 945, 953–54 (1990); Matthew D. Adler, *Linguistic Meaning, Nonlinguistic "Expression," and the Multiple Variants of Expressivism: A Reply to Professors Anderson and Pildes*, 148 U Pa L Rev 1577 (2000).

[22] Grice, *Meaning* at 436 (cited in note 20).

[23] Id at 438.

ings, both because there are as many natural meanings of a given crime as there are causes (making it entirely accurate to say that "the murder meant a blood sugar shortage"), and because the natural meaning of a crime would never have false propositional content that invites or demands contradiction. Only if crimes can "say" false things can they—and perhaps, then, should they—be contradicted. Inasmuch as effects cannot lie about their causes—and a crime has the natural meanings that it does only by virtue of the causes of which it is, in fact, an effect—they cannot be construed as false propositions in need of correction.

This is to say, then, that expressivists must construe crimes as communications of a sort—as "signals" possessed of non-natural meaning, or "meaning$_{NN}$,"[24] rather than as "signs" possessed of natural meaning. Only if crimes are construed as intentional human communications can they possess propositional content of questionable truth value that would merit efforts of falsification by others. We are thus naturally compelled to turn to classic communication theory in order to make sense of how a crime could constitute a communication possessing Gricean "meaning$_{NN}$."

On the standard view, communication requires what J.L. Austin referred to as an "illocutionary speech act."[25] An illocutionary speech act is performed whenever one utters a sentence or performs an action that has particular "conventional force" indicating what one is intending to convey by one's utterance or action.[26] This communication typically involves three elements.[27]

First, one who communicates means something by his action—his action has what linguists call speaker's, or utterer's,

---

[24] Id at 437.

[25] J.L. Austin, *How to Do Things with Words* 109 (Harvard 1975).

[26] Id at 103, 109. Illocutionary acts are performances of acts *in* saying something. Id at 99. Telling someone that her house is on fire, for example, is an illocutionary act because A is warning B of danger. Id at 103–04. So, for example, I can say "I advise you to go to the store," but I cannot say, "I convince you to go to the store" (because the latter is a "perlocutionary" act that requires participation from B, in which the performance of the act is done *by* saying it). Id at 101–04.

[27] For a classic summary of the elements of communication, see Grice, *Meaning* at 441 (cited in note 20). For an appreciation of the extent to which Grice's analysis of the conditions of communication has been accepted as definitive by those who work in communication theory and the philosophy of language, see Mark Platts, *Ways of Meaning: An Introduction to a Philosophy of Language* 86–88 (Routledge & Kegan Paul 1979); Robert Martin, *The Meaning of Language* 83–95 (MIT 1987); Peter Strawson, *Logico-Linguistic Papers* 155–58 (Metheun 1971); John Searle, *What is a Speech Act?*, in Rosenberg and Travis, eds, *Readings in the Philosophy of Language* at 620 (cited in note 20).

meaning.[28] Second, a rational communicator will typically choose conventional means of expressing his message. In linguistics, such conventions are often (if controversially) divided into semantic conventions (giving the meaning of words, symbols, and sentences) and pragmatic conventions (giving the norms of appropriate utterance). Finally, when the beliefs of the audience match the intentions of the speaker—so as to achieve what J.L. Austin called audience "uptake"[29]—then communication has succeeded.[30] If crimes can be understood communicatively, then they should adhere, at least roughly, to this model of communication.

It should be clear from this briefest of surveys that the expressivist's claim that crimes convey messages cannot be vindicated by easily applying the standard model of communication. For it seems obvious that criminal deeds are generally not illocutionary acts: they do not represent conventional means by which speakers seek ways of conveying particular messages. In the tripartite sense in which communication theorists think of communications, crimes fail to satisfy the first criterion, for they fail to possess "speaker's meaning."

Some criminal actions, of course, *are* calculated to send a message to their victims. Certain hate crimes, for example, might be thought to be strategically designed to be substitutes for raw threats: they are perpetrated with the intent to communicate to victims clear messages of contempt and clear warnings concerning their long-term safety within the community.[31] But far more often, crimes have no such illocutionary intentions behind them. Murderers often kill and rapists often rape without intending that their victims achieve audience uptake of any sort. Their goals are to kill and rape, not to talk. Their crimes cannot be thought to have "speaker's meaning" because there is nothing that they mean *by* them.

The next move open to the expressivist is to invoke the third element of a full-fledged theory of communication—the element

---

[28] The concept of speaker's meaning again finds its basis in the work of H.P. Grice. See Grice, *Meaning* at 386 (cited in note 20) (describing the "utterer's intention"). J.L. Austin called an actor's meaning the actor's "locutionary act." Austin, *How to Do Things with Words* at 94 (cited in note 25). And Michael Moore has termed it the speaker's "semantic intention." Michael S. Moore, *A Natural Law Theory of Interpretation*, 58 S Cal L Rev 277, 340–44 (1985).

[29] Audience "uptake" occurs when the audience understands the meaning and the force of the locution. Austin, *How to Do Things with Words* at 116–17 (cited in note 25).

[30] Heidi M. Hurd and Michael S. Moore, *Punishing Hatred and Prejudice*, 56 Stan L Rev 1081, 1102 (2004). Moore, 58 S Cal L Rev at 340–44 (cited in note 28).

[31] Dan M. Kahan, *What Do Alternative Sanctions Mean?*, 63 U Chi L Rev 591, 598 (1996).

of audience belief. As the argument would go, crimes have meaning to the extent that those who are audiences to them attribute meaning to them. Thus, if an audience takes a grisly murder to possess meaning—attributing to it, for example, the proposition that the victim's life was of no worth—then the crime has meaning$_{NN}$. And if the proposition for which the audience takes the crime to stand is deemed false, then the expressivist could claim that punishment should be imposed in a manner that will cause that same audience to reach a contradictory judgment from that proffered by the crime.

It is far from nonsensical to attribute a meaning to an action solely on the basis of how an audience understands that action. Audience-oriented interpretation is well established and makes good sense in several legal contexts, including defining what counts as defamation,[32] establishing a violation of the Lanham Trade-Mark Act,[33] or fixing the meaning of a contract. Thus, for example, whether a defendant's statement is slanderous to a plaintiff turns on whether at least a significant minority of the community (a "*Peck* audience"[34]), if not a majority, takes it to be of and concerning the plaintiff and is inclined to subject the plaintiff to ridicule and contempt as a result of it.[35] That the defendant did not intend the statement to be about the plaintiff or to be besmirching to reputation is neither here nor there, provided that others take its reference to pick out the plaintiff and take it to be derogatory.

There are, however, two problems with invoking defamation law's concept of audience understanding in order to fix the social meaning of a crime. First, how should we fix the audience to a criminal's actions so as to extract the social meaning of those ac-

---

[32] See Restatement (Second) of Torts § 559 (1965) ("A communication is defamatory if it tends so to harm the reputation of another as to lower him in the estimation of the community."); *West v Thomson Newspapers*, 872 P2d 999, 1008 (Utah 1994) (stating that the guiding principle in whether a statement is defamatory, aside from its truth, "is the statement's tendency to injure a reputation in the eyes of its audience").

[33] 15 USC §§ 1041–1127 (2000). The Lanham Trade-Mark Act authorizes the Director of the Patent and Trademark Office to deny a trademark registration if she determines that the public is likely to confuse it with another registered trademark. 15 USC § 1052(d). The Lanham Act also creates a civil cause of action if a person uses in commerce any misleading description of fact which is likely to deceive, cause confusion, or cause mistake in regards to affiliation with another person's brand of merchandise. 15 USC § 1125(a).

[34] *Peck v Tribune Co*, 214 U.S. 185 (1909).

[35] Id at 189–90 (holding that "general consensus" is not necessary to prove libel; instead, plaintiff must only prove that the alleged libel "will be known by a large number and will lead an appreciable fraction of that number to regard the plaintiff with contempt").

tions from the audience's understanding of them? It will not do to adopt a majoritarian measure, unless we are truly prepared to say that slavery is not derogatory so long as the majority does not think of it as such. And if we are prepared to say that an action has a given social meaning if a minority would take it as such, then many criminal deeds may have two or more social meanings of contradictory propositional content. The social meaning, for example, of the gang member's crime may be deemed laudatory by his gang, while being perceived as an expression of contempt by the victim's community. Which meaning are we to privilege?

Second, it seems close to impossible to attribute any discrete meaning to a given crime that distinguishes it in its meaning—and so in its punishment—from other crimes. Just try it. Imagine that a drunk driver takes the outside corner of a blind curve, but avoids hitting anyone. You are an audience to his recklessness. What (false?) proposition do you attribute to his deed? And how does that proposition differ from those propositions that you assign to car thefts, muggings, tax evasion, speeding offenses, or acts of petty vandalism? If these crimes do not possess quite different and quite specific meanings in your mind—if you cannot attribute to them particular propositions concerning their victims—then in what sense could you craft specific punishments that would contradict, with precision, the terms of the messages sent by these crimes?

As I have tried to craft examples that will sympathetically illustrate the expressivist's thesis within this paper, I have found myself repeatedly throwing up my proverbial hands and simply describing a particular crime as conveying "disrespect for the victim." But unless a rape is properly punished the same as a petty theft (and vice versa)—because in the end both crimes convey an identical message of disrespect for their victims—the notion that crimes possess distinct meanings by virtue of *Peck* audiences assigning to them distinct propositions is more metaphorical than meaningful.

The third possibility that the expressivist might pursue in making sense of the social meanings of crimes—and thus punishments—is to take the conventions that make communication possible as the locus of meaning. After all, communication between persons reflects a very significant coordination problem that can be, and has been, solved only by the development of shared conventions. Consider the familiar example of a novel typed by the 500,000th monkey in the British Museum whose random tapping of keys miraculously produces not only syntacti-

cally correct sentences but also strings of them with theme, narrative consistency, interesting plot development, and so on. The product of the monkey's random typewriter tapping has meaning even if we must admit that no one wrote it and even if no one reads it. Its meaning derives from its conformity to the semantic, pragmatic, and stylistic conventions that attend novelistic utterances in English.

It might be argued then, that the social meaning of crimes is given by whatever conventions they reflect—regardless of whether the offender intends to invoke those conventions, and regardless of whether the victim—or anyone else—in fact perceives the message accordingly. An action is racist if it invokes norms that have historically been employed to communicate messages of bigotry; an action is chauvinistic if it participates in conventions that have been historically employed to reinforce gender distinctions; and so forth. Thus, even if a defendant had no knowledge of the conventions associated with spray-painting a swastika-like symbol on a synagogue, his act would have anti-Semitic meaning, because conventions alone give it that meaning—even if no one experiences it as anti-Semitic, and despite its not being intended as such.

The problem with an effort to derive social meaning from the second element of the standard account of communications is that conventions are famously under- and over-determinate. What is the conventional meaning, for example, of a man's pulling out a chair for a woman? Does it honor her or disrespect her? Since the act simultaneously participates in or draws upon a multitude of conventions—some honorific, some oppressive—it is extraordinarily hard to fix its social meaning through these social conventions. While a swastika has quite clear connotations, few crimes employ symbols that have discrete meanings within our language and culture. What particular, discrete conventions are invoked by or reflected in drunk driving, illegal prostitution, mail fraud, arson, and public exposure such that they can be employed to give these crimes meanings distinct from one another and meanings that could be contradicted by particular punishments?

These problems, it seems to me, are sufficient to suggest that expressivists need a theory of social meaning that divorces meaning from the illocutionary intentions of offenders—that is, from speakers' intentions; the beliefs of victims or witnesses—that is, from audience beliefs; and social norms—that is, from

semantic and pragmatic conventions. As I have discussed in substantial detail elsewhere,[36] many expressivists remarkably seem to think that they can build such a theory without any of these theoretical blocks. In a leading article on legal expressivism Elizabeth Anderson and Richard Pildes tell us that public meaning "need not be in the agent's head, the recipient's head, or even in the heads of the general public."[37] They offer the following example: "Musicians can play music that expresses sadness, without feeling sad themselves. The music they play need not express their (or anyone's) sadness: the sadness is in the music itself."[38] It is surely true that music can express sadness without the composer being sad at the time that she wrote it, and without the musicians themselves feeling sad at the time that they play it. It can also express sadness without an audience on a given occasion being caused either to feel sad or to believe that the music is sad. But can music express sadness if the music conforms to no musical conventions associated with the conveyance of a sad mood, so that it has no propensity whatsoever to elicit these cognitive or emotional experiences in audiences? Without actual mental states on the part of authors or audience members, and without any norms abstracted from other authors and other audiences reacting to different, but relevantly similar, kinds of music, in what possible sense could music *be* sad or *convey* or *express* sadness?[39]

Let us consider three possibilities. The first has been suggested by Simon Blackburn in his efforts to ascribe some sensible meaning to the expressivists' notion of social meaning.[40] Blackburn offers expressivists what he calls a "credibility" construal: an action A expresses some meaning M when M is the only credible motivation that an actor could rationally have had in doing A.[41] Thus, to use one of Blackburn's examples, the flying of the Confederate flag over the government buildings in certain

---

[36] Hurd and Moore, 56 Stan L Rev at 1106–10 (cited in note 30).

[37] Elizabeth S. Anderson and Richard H. Pildes, *Expressive Theories of Law: A General Restatement*, 148 U Pa L Rev 1503, 1525 (2000).

[38] Id at 1508.

[39] See Steven D. Smith, *Expressivist Jurisprudence and the Depletion of Meaning*, 60 Md L Rev 506, 562–63 (2001) (arguing that if the music in Anderson and Pildes's example does not express sadness felt by anyone, then it makes no sense to call it "sad").

[40] Simon Blackburn, *Group Minds and Expressive Harm*, 60 Md L Rev 467 (2001). For a somewhat more extensive discussion of Blackburn's proposal, see Hurd and Moore, 56 Stan L Rev at 1107–08 (cited in note 30).

[41] Blackburn, 60 Md L Rev at 483 (cited in note 40). Blackburn sometimes puts his credibility account slightly differently: An action A expresses M when there is no credible way that the actor could rationally engage in A without defending M.

Southern states can be thought to expresses racism if the only credible reason that a rational agent could have for flying such a flag is to celebrate the racial hierarchy that characterized the South in pre-Civil War days.[42]

Our problem, again, will be that the credibility condition appears likely to over- or under-determine the social meaning(s) of an action. Whether it can uniquely determine the (rational) motive(s) to be attributed to an action depends on how stringent the norms of rationality are taken to be. If one takes them to be very stringent, then there may be but one rational end for a given action. If one takes them to be very loose, however, then there may be many possible motives for any action. Blackburn presumably would intend the expressivist to couple his credibility construal with a level of stringency that is taken to match the average information base, average inference-drawing capacities, and average intelligibility of desires of humankind.

Blackburn's construal of "social meaning" holds out the promise of giving the concept sufficient content that it avoids what Blackburn rightfully fears for it, namely, that "anything goes"[43]—that is, that any meaning can be assigned to any crime so as to justify any punishment. The solution that Blackburn offers the expressivist, however, would make a defendant deserving of punishment not because of any fact about him or his deed, but because of the *appearance* of some fact about him or his deed: namely the appearance that he possessed a particular motivation for his crime, regardless of whether he in fact possessed such a motivation and regardless of whether anyone perceived him as having such a motivation. To declare that a crime means what a hypothetical rational actor would have meant by it had she performed it under the circumstances—and to punish a defendant so as to contradict that message—is to punish someone for the hy-

---

[42] Id at 485. It is tempting to regard Blackburn's credibility construal of social meaning as but another return to the subjective mental states of actors, the only novel twist being a stringent evidentiary requirement for when an actor can be thought to possess a mental state. Yet Blackburn plainly intends no such construal, for his aim is to provide a notion of social meaning that can apply to groups and institutions that in no obvious sense actually possess mental states. Blackburn's is a constructivist construal of social meaning, one that is based on the mere appearance, and not the reality, of there being a certain motivation. Such an appearance is generated by attributing rational beliefs and rational inferences to the agent (or entity) whose act is thought to have social meaning and asking (in light of the causal and other properties of the action, and in light of the absence of any better means to achieve a certain state of affairs) what end could most rationally have been achieved by such a means. Hurd and Moore, 56 Stan L Rev at 1107 n 68 (cited in note 30).

[43] Blackburn, 60 Md L Rev at 478–79 (cited in note 40).

pothetical and not for the real. If expressivists can live with this result, then their theory likely has normative problems that will swamp its conceptual problems.

The second construal of the concept of social meaning that does not rely on authorial intentions, audience beliefs, or existing communicative conventions is offered by Deborah Hellman in her defense of an expressivist approach to equal protection analysis in constitutional law.[44] Hellman appeals to audience beliefs in her construal, but they are not the actual beliefs of crime victims, witnesses, or consumers of current events. Rather, in her view, we should decipher the hypothetical beliefs that a citizenry would have if it were to engage in dialogue in a Habermasian "ideal speech" situation.[45] As she puts it, "the expressive dimension of a law or policy is best understood as the meaning that we would arrive at if we were to discuss the interpretive question together under fair conditions."[46]

It is unclear how an expressivist who seeks to make use of this suggestion in the context of justifying criminal sanctions would flesh out this claim. How would we fix the "fair conditions" for a conversation about the "interpretive question" of the social meaning of a criminal's action? Specifying such conditions, however, is not the least of the problems that confront Hellman's approach. The conclusion of such an idealized epistemic exercise cannot even claim the relevance of an *appearance* of a bad motivation. Such a conclusion can yield only the judgment that there *would be* such an appearance in the idealized situation imagined. It is morally unintelligible how one could conclude that the state would be justified in punishing those whose crimes *would be* viewed as sexist or racist by hypothetical people engaged in certain idealized dialogues in another hypothetical world.

A third possible unpacking of the expressivist's concept of social meaning—one that does not rely on any of the standard components of communication theory—is offered by another

---

[44] Deborah Hellman, *The Expressive Dimension of Equal Protection*, 85 Minn L Rev 1 (2000).

[45] Id. See Jurgen Habermas, *Justification and Application: Remarks on Discourse Ethics* 54–60 (MIT 1993) (Ciaran Cronin, trans) (arguing that the ideal speech situation must include freedom of access, equal rights to participate, truthfulness on the part of participants, absence of coercion in taking positions, and similar values). For an expanded treatment of the ideal speech situation, see Jurgen Habermas, *Moral Consciousness and Communicative Action* 83–94 (MIT 1990) (Christian Lenhardt and Shierry Weber Nicholson, trans).

[46] Hellman, 85 Minn L Rev at 23 (cited in note 44).

critic of expressivism, Steven Smith.[47] Smith also engages in an extensive search for a conception of social meaning that is coherent and can do the normative work that it is assigned by expressivists.[48] In exasperation, Smith concludes that "expressivist scholars seem to suggest that laws or actions can just have 'objective' meanings, period—meanings that need not be meanings of or to *anyone*."[49] Smith concludes that "this assertion simply renders the notion of 'meaning' unintelligible."[50]

I agree with Smith's conclusion that those with expressivist intuitions ultimately believe that social meaning is a primitive or raw concept that is unresponsive to analysis. Certainly those who wish to free themselves from the constraining implications of the various theories concerning what social meanings can be—theories framed in terms of intentions, beliefs, or conventions—do seem to yearn for an account of meaning that gives it metaphysical autonomy. The most flagrant example of one who espouses this metaphysically primitive view of meaning is Edwin Baker.[51] According to Baker, there is a "missing realm" of existence that, once recognized, solves the problems that have concerned us.[52] On his understanding, there is both the "material realm" of the natural world and the "subjective realm" of the mind; but there is, in addition, a "third realm": the "social realm."[53] Discovery of this realm solves the conceptual worry concerning the locus of social meaning because "meaning exist[s] here, in this social realm, not as [a] 'mental event[]' in the heads of either creators or perceivers."[54] And this construal of social meaning purportedly solves the normative worries that I have registered in response to some attempts to fix social meaning, because according to Baker, "a proper normative concern focuses on 'meanings' that exist in the social realm."[55] Baker concludes that "an understanding of this 'social' realm is absolutely

---

[47] Smith, 60 Md L Rev 506 (cited in note 39).

[48] Id at 510–18.

[49] Id at 562.

[50] Id at 563.

[51] See C. Edwin Baker, *Injustice and the Normative Nature of Meaning*, 60 Md L Rev 578 (2001) (detailing the merits of an expressivist jurisprudence that would prohibit the government from acting in a manner that expresses a constitutionally forbidden meaning). For a similar discussion of Baker's free-standing conception of meaning, see Hurd and Moore, 56 Stan L Rev 1109–10 (cited in note 30).

[52] Id at 583.

[53] Id at 584–85.

[54] Id at 586.

[55] Baker, 60 Md L Rev at 585 (cited in note 51).

fundamental to an understanding not only of meaning and interpretation, but of the social sciences generally."[56]

But as one critic of metaphysical promiscuity has argued, "special realms" are the last refuges of theories in trouble.[57] One who postulates metaphysical divides between the natural and the moral, between the physical and the mental, or between the individual and the social commits himself to what John Mackie called "queer entities."[58] And as I have suggested elsewhere,[59] once one commits oneself to metaphysical queerness, one is surely committed to postulating queer relations between queer and non-queer entities, and queer epistemological means of gaining knowledge about such entities and their relations. And even if we ignore all of these vintage objections to the kind of pluralist metaphysics upon which Baker's view relies, we are returned to Blackburn's fear that we have entered a "hermeneutic desert,"[60] a landscape "where anything goes."[61]

It would seem, then, that expressivism cannot deliver up a concept of social meaning that is coherent and determinate in its implications for particular cases. Absent a theory of how to attribute meanings to crimes that are propositionally unique and therefore capable of being "contradicted" by particular penalties, expressivism will be unable to make sense of when and why particular offenders ought to receive particular punishments.

Let us now leave the conceptual problems that confound attempts to make sense of the expressions that expressivists attribute to crimes and punishments and turn to two normative questions that must be answered by those who take expressivism to be an able contender within punishment theory.

B. Making Moral Sense of Using Punishments to "Contradict" the Social Meaning of Crimes

Expressivists must think that it is good for society to express its disapproval of crimes—indeed, so good as to justify the intentional infliction of the suffering that is punishment. So the first normative question that expressivists must confront is why one should think this. There are some easy answers that are *not*

---

[56] Id at 584 n 9.

[57] See Michael S. Moore, *Educating Oneself in Public* 384–87 (Oxford 2000).

[58] J.L. Mackie, *Ethics: Inventing Right and Wrong* 38–42 (Penguin 1977).

[59] Hurd and Moore, 56 Stan L Rev at 1110 (cited in note 30).

[60] Blackburn, 60 Md L Rev at 482 (cited in note 40).

[61] Id at 479.

available to expressivists. These are the answers that draw on a utilitarian theory of punishment—a theory that takes the expression of societal outrage to be an *instrumental* good in the service of other, intrinsic goods.

As I outlined in Part I-C, a utilitarian might plausibly suppose that the denunciation of crimes will allow a society to vent its vengeful emotions in a way that prevents vigilante justice. She might believe that such denunciation will reinforce the values of law-abiding citizens in a way that keeps them law-abiding. She might believe that denouncing crimes will frighten others who are less law-abiding or educate those who are ignorant. She might simply believe that such denouncement satisfies the wants of most people, or provides them with a vital means of social cohesion and shared identity.

The expressivist, however, cannot argue that expressions of social condemnation will yield this familiar set of goods without reducing expressivism to a mere means of achieving a utilitarian program. He cannot argue that we should employ punishment to express our highest and best ideals if the reasons that it is good to do so depend upon other goods. For the expressivist promised more. He promised us a theory of punishment independent of the traditional theories of punishment, and thus independent of claims that it will be an effective means of advancing the utilitarian's agenda. If expressivists are to deliver on this promise, then the expression of disapproval by criminal punishment must be an *intrinsic* good, not merely an instrumental good in the service of utilitarian goals.

For like reasons, expressivists cannot cobble their claims to those of retributivists by arguing that the expressive function of law comes into its own when the demands of desert are vague. They thus must reject Fred Lawrence's claim that the expressive effect of a given punishment must at least "inform our decisions about the nature of that punishment," even if it does not, by itself, justify the punishment.[62] Nor can expressivists urge, as Dan Kahan does, that "[t]he proper retributive punishment is the one that appropriately expresses condemnation and reaffirms the values that the wrongdoer denies."[63] If according offenders their

---

[62] Frederick M. Lawrence, *Punishing Hate* 167 (Harvard 1999). See also Kahan, 63 U Chi L Rev at 602 (cited in note 31) (noting that the expressivist view informs desert when an individual "engages in behavior that conveys disrespect for important values.").

[63] Kahan, 63 U Chi L Rev at 591 (cited in note 31). For a more lengthy discussion of the incompatibility of expressivism and retributivism, see Hurd and Moore, 56 Stan L Rev at 1112–13 (cited in note 30).

just deserts justifies punishing them at all, then it justifies both the kind and quantity of punishment an offender should receive. There is no conceptual room for a principle at odds with retributivism to do work *within* a retributivist program. And *even if* ambiguities inherent in a retributivst theory required recourse to a foreign factor—say, expression of social outrage—then that factor would be doing all the justifying work with regard to the matters that are considered ambiguous. And then the normative question with which we began would recur: What makes expressing a society's disapproval of an offender's criminal conduct intrinsically good? *Ex hypothesi*, it cannot be that it is deserved. What accounts for its value?

Expressivists might be tempted to argue that if a good is intrinsically good, then no argument is possible either for or against it. An intrinsic good is, as it were, another primitive. Yet the sheer fact that something is intrinsically good does not mean that one cannot advance reasons to conclude that it has intrinsic goodness. Certainly one cannot show it to be intrinsically good by showing how it advances something else that is good. But this is similarly true of categorical obligations, which cannot be justified by reference to good consequences. Deontological permissions and prohibitions can be justified by demonstrating how they cohere with a mix of particular judgments and general principles that one accepts as provisionally true.[64] It would simply be false to claim that nothing can (or need) be said to justify a belief in the intrinsic goodness of an act, or an institution, or a practice. So the expressivist cannot escape the question of how he is to defend, as inherently good, the expressive use of the power of punishment.

In the end, it seems to me that there is no plausible answer to this question. Some institutions and institutional practices are arguably intrinsically good, such as educational institutions that come as close as possible to affording equality of opportunity. But employing the criminal law so as to express popular attitudes does not seem of similar inherent worth. Those who consider the expression of social attitudes to be intrinsically good, as opposed to being good for something, might have a sense that there is a kind of collective First Amendment right; a right justified by the intrinsic goodness of "self"-expression. But when the self is not

---

[64] For an insightful discussion of how one justifies first principles and how such principles differ from intrinsic goods, see Michael S. Moore, *Placing Blame: A General Theory of the Criminal Law* 159–87 (Clarendon 1997).

an individual, the claim that its expression is an intrinsic good becomes bizarre. It seems a kind of category mistake to liken the nature and value of the collective expression behind a criminal punishment scheme to the nature and value of an individual's expression in a work of art or piece of poetry.

C. Making Moral Sense of Achieving Societal Expression Through Undeserved Punishment

Let us suppose that we find some basis for concluding that sending denunciatory messages to criminals via punishment is intrinsically (as opposed to instrumentally) good. The next question is this: Can sending a contradictory message be so good that it justifies punishing an offender more than is deserved? This question arises because we are to assume that expressivism can justify punishment even when desert-oriented retributivist and mixed theories cannot. By its own claim, it constitutes a separate and autonomous account of when and why punishment is justified. Hence, it must be the case that, at least in principle, it can sometimes license punishment beyond what would be deserved. Consequently, we must ask whether the expressivist can plausibly claim that the good inherent in social condemnation of crime has the capacity to justify punishing a criminal more than she deserves.

If I were an expressivist about punishment, I would resist the question. I would argue that the obligation not to punish an offender beyond what he deserves is internal to the good of expression. That is, the good of expressing condemnation is a good at all only when the criminal being condemned fully deserves the punishment that expresses the condemnation. Because this is an intrinsic good, it is not possible to achieve such a good absent desert. The obligation not to punish an offender beyond what he deserves would be, on this argument, not a side-constraint to attaining the good of expression; it would be internal to that good. One could, I would argue, achieve the educational benefits and other utilitarian gains caused by expressions of condemnation even when the person who is condemned is innocent, but we have already conceded that these benefits are not what make the expression of condemnation *intrinsically* good.

An expressivist who seized my argument as a friendly amendment would certainly have to construe his theory of punishment as deontological in character. He would have to argue that we are categorically obligated not to express

condemnation of those who do not deserve it, even when such condemnation would maximize the expression of condemnation of those who do deserve it. Such an argument would insulate the expressivist's theory from the charge classically made against the utilitarian: namely that his theory permits the punishment of an innocent man.[65]

But mine is a false friendship. One cannot accept my offer to rescue expressivism from the perils of utilitarianism without completely trivializing the theory. For my amendment makes the expressivist theory of punishment incapable of justifying any punishment beyond what is deserved. So it seems that the expressivist must bite the same bullet that has been rotting the teeth of utilitarians over the years, for he must be prepared to "communicate" with the public via the imposition of undeserved punishment if he believes in his own theory and its autonomy from other theories.

A final and related normative challenge is this one: Can sending a message be so good that it justifies punishing an offender less than is deserved? In many cases, it would seem that we could surely express our condemnation of criminals without actually punishing them.[66] That is, we could obtain the good sought by the expressivist theory—we could denounce what deserves denouncing and contradict what demands contradiction–without in any other way making those who are contradicted and denounced suffer. If true, then the expressivist, like the utilitarian, must admit that when condemnation can better be achieved through means other than punishment, desert drops out as a condition of punishment. That an offender richly deserves to be punished is neither here nor there if he and his act can be adequately condemned without such punishment.

Expressivists will surely handle the prospect of under-punishment by repairing, in the words of one, to "what convention and form contribute to meaning."[67] Recognizing that "[i]n some societies, and even in ours at an earlier time, public denunciation by itself might have been sufficient to convey condemnation of a wrongdoer," Kahan claims that in our society today, only the imposition of suffering on an offender can express

---

[65] See Part I-C.

[66] See H.L.A. Hart, *Law, Liberty and Morality* 66 (Stanford 1963) (arguing that "the normal way in which moral condemnation is expressed is by *words*" and not by punishment) (emphasis in original).

[67] Kahan, 63 U Chi L Rev at 600 (cited in note 31).

our condemnation of him and his action.[68] "[T]he way for society to show that it takes rape seriously, and to show that it genuinely condemns a particular rapist, is to make him suffer."[69]

But such a refusal of the question will not do. For even if Kahan has the sociology right (as he may well have), such a response hardly meets the challenge that has been posed. The question is not whether the expressivist can anticipate that in present day circumstances his theory will permit the deserving to go free. The question is whether he can stomach his theory when happenstance makes it less happy. Can the expressivist accept the implications of his theory in the case of a brutal rapist who fully deserves punishment but who is lucky enough to live in a society in which verbal condemnation is understood as sincere and sufficient to convey due respect for the victim?

Even if expressivists succeed in characterizing as de minimus the concern that contradiction and condemnation might someday take a form quite different from punishment, they would continue to be faced with the same prospect as that of utilitarianism—that a society might (and ought to) merely *pretend* to punish whenever such a pretense can be carried out without detection. Why not create a brutal-rapist protection program that gives rapists new identities in a faraway place, while using the conventions of condemnation that exist in our society to make a great show of (seemingly) punishing such long-gone offenders? The expressivist can always insist that there is an ineliminable risk that such false shows of punishment will be discovered. But such a risk ought only to give the expressivist a discount rate. On his own theory, if it is good to achieve public condemnation of the social meaning of crime and bad to make individuals suffer, then he should take some risks whenever he has a good shot at achieving the good without the bad. And nothing beyond concern for detection should give the expressivist pause when pretense can do what he would otherwise ask of punishment.

It would thus seem that expressivists fare no better than corrective justice theorists, rehabilitation theorists, utilitarians, and mixed theorists in their ability to escape the charge that they have unjustifiably elevated their normative ideal over that of achieving punishments that match just deserts. They must thus join such other more traditional theorists in the

---

[68] Id at 600–01.

[69] Id.

philosophical effort to persuade laypersons and academics alike that there is no moral magic in the claim that persons should receive only what they deserve—a task that I take to be both unpromising and unworthy.

CONCLUSION

I have argued that there is good reason for criminal law theorists to be unsatisfied with the menu of punishment theories that has long specified the philosophical fare in criminal law. As one canvasses the traditional entries, as I did in Part I, one is struck by the seriousness of their conceptual and normative problems. One well understands how there could be a market for a fresh alternative—one that holds out the promise of providing a normatively compelling reason to punish culpable offenders without the conceptual problems that confound the corrective justice theory, rehabilitation theory, utilitarian theory, mixed theory, and retributive theory.

As I have argued, however, the newest temptation—expressivism—is to be resisted.[70] While some of the academy's most creative contemporary thinkers have been captured by the claim that punishment can be justified by its ability to contradict and condemn the social meaning of crimes, I have outlined here why one ought to have grave doubts that expressivism can lay claim to normative superiority over desert-based theories of punishment. I have further shown why it invites far more serious conceptual challenges than do any of the traditional philosophical perspectives, given its reliance on the amorphous (and probably incoherent) notion of crimes having something called "social meaning" that can be "contradicted" by punishments. So I am going to return to toiling in traditional fields, leaving those who have a fondness for bandwagons to

---

[70] Even Joel Feinberg, whose early article *The Expressive Function of Punishment* is often cited by present-day expressivists, always held that expression of condemnation is at most an incidental function of punishment and is not part of a properly conceived theory of punishment, namely a theory that justifies the institution. Joel Feinberg, *The Expressive Function of Punishment*, in *Doing and Deserving: Essays in the Theory of Responsibility* 95, 95–98 (Princeton 1970) (arguing that punishment's expressive function of conveying society's condemnation of an act is what distinguishes "punishments" from mere "penalties"). See also Letter from Joel Feinberg, Professor, Arizona State University, to Michael Moore, Professor, University of Illinois (Apr 1990) (on file with author).

what I take to be the hopeless task of fixing expressivism's broken wheels.