Matthew C. Altman
Editor

The Palgrave Handbook on the Philosophy of Punishment

palgrave
macmillan

Editor
Matthew C. Altman
Philosophy & Religious Studies
Central Washington University
ELLENSBURG, WA, USA

ISSN 2730-9169 ISSN 2730-9177 (electronic)
Palgrave Handbooks in the Philosophy of Law
ISBN 978-3-031-11873-9 ISBN 978-3-031-11874-6 (eBook)
https://doi.org/10.1007/978-3-031-11874-6

Cover illustration: Library of Congress, Prints & Photographs Division, photograph by Carol M. Highsmith [reproduction number, e.g., LC-USZ62-123456]

This Palgrave Macmillan imprint is published by the registered company Springer Nature Switzerland AG.
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

4

Is Crime Caused by Illness, Immorality, or Injustice? Theories of Punishment in the Twentieth and Early Twenty-First Centuries

Amelia M. Wirts

In the twenty-first century in the United States, discussions of punishment often raise critical questions about mass incarceration, racial and class disparities in the criminal justice system, mental health crises among incarcerated and poor communities, and the moral legitimacy of policing and incarceration. These popular discussions implicate philosophical questions that are perennial. Is the purpose of punishment to bring about social benefits such as reducing crime, rehabilitating those who commit crimes, and protecting the public from threats to safety? Those who think that crime primarily occurs because those who commit crimes lack mental health treatment or the right incentives to follow the law will likely take this *utilitarian* view of punishment. Or, is punishment always the right response to crime, regardless of whether punishment brings about any social benefits? Those who think that crime is a moral problem that stems from the moral failings of individuals as agents will likely take this *retributivist* view of punishment. For retributivists, we punish because imposing hardship on those who commit crimes is simply the right thing to do. These two approaches to thinking about the causes of crime and the proper punitive responses animated debates in punishment theory in the twentieth century.

A. M. Wirts (✉)
University of Washington, Seattle, WA, USA
e-mail: amwirts@uw.edu

M. C. Altman (ed.), *The Palgrave Handbook on the Philosophy of Punishment*, Palgrave Handbooks in the Philosophy of Law, https://doi.org/10.1007/978-3-031-11874-6_4

In the early twentieth century, punishment theory was dominated by the rehabilitative ideal. This particular strand of utilitarianism presented much like a medical model of crime. Crime was like an illness, and punishment was best understood as treatment for the person who committed a crime. Punishment helped the person recover from the illness of criminality by offering mental health care, treatment for addiction, job training, or other social services.

In the 1970s, the consensus was that the rehabilitative model failed. This may have been fueled by increasing empirical evidence that rehabilitation did not actually help people commit fewer crimes, or by rising crime rates, or both. In its wake, philosophers took up retributivist justifications of punishment. Retributivists railed against the view that criminals were ill, arguing that this undermined the agency of those who committed crimes. The rehabilitative model, they argued, misunderstood the problem of crime. It was a moral problem, not a health problem.

Utilitarian and retributivist approaches to theorizing crime and punishment have dominated philosophy of punishment since the beginning of the twentieth century (Feinberg 2008). But there was also another view of crime that motivated social movements against police brutality and prisons during that time, even if it was more rarely discussed in philosophy journals or law schools. On this view, crime was a consequence of an unjust social structure, not of psychological or moral deficiencies of individuals. The *social critical* view of crime entailed that punishment cannot address the underlying issues of crime because punishment can only address the individual, not the society. While the classical debates in philosophy of punishment were framed around what can justify punishment practices, both these theories relied, implicitly or explicitly, on the idea that crime stems from defects in the individual person. In contrast, social critics of punishment have instead argued that crime stems from moral defects in society itself, including racism, poverty, and ablism.

Increasingly, the social critical view has come to animate public protests and popular discourse about policing and mass incarceration, especially since 2020. Social critics continue to argue that the practice of punishment cannot address the problem of crime because punishments, by their nature, address individuals. For example, Angela Davis (2003) has consistently argued that crime is an excuse to maintain anti-Black racial hierarchies, not to keep society safe. She also argues that the way to address crimes is to understand them as effects of social injustice rather than individual moral failings or disorders.

This chapter makes the case that, since 1900, debates about the justification of punishment have also been debates about the cause of crime. In part one, I explain how the rehabilitative ideal of punishment viewed mental illness and dysfunction in individuals as the cause of crime. Since rehabilitative models found social and mental defects in the individual as causes of crime, they treated crime with mental health care and social training. In section 2, I argue that mixed models of punishment criticized the rehabilitative view that most people who commit crimes lack agency, but mixed models still maintained the view that mental illnesses and other circumstances could radically undermine a person's agency. H. L. A. Hart's criticisms of rehabilitation presaged those of the retributivists. In section 3, I argue that retributivism was best understood as identifying the immorality of human agents as the source of crime, which dovetailed well with the "tough-on-crime" political milieu of the 1980s and 1990s that produced mass incarceration. The only legitimate response to crime was a kind of punishment that addressed the moral failings of the person. In section 4, I briefly offer an alternative to both retributivism and rehabilitation, which both found deficiencies in individuals to be the cause of crime. Following Davis, I suggest that crime is best understood as a product of an unjust society, not faulty human beings. Thus, punishment, which is only aimed at individuals, cannot address the deeper causes of crime. But, since this view tends to emphasize rebuilding social services, it must take the lessons learned by the critiques of rehabilitation and resist the tendency to reduce human beings to recipients of those social services.

1 The Early Twentieth Century: The Reign of Rehabilitation and Incapacitation

In the first half of the twentieth century, philosophers, criminologists, and legal institutions emphasized rehabilitation and incapacitation (von Hirsch 1985; M. Davis 1990, 2009). According to legal scholar Francis Allen:

> The rehabilitative ideal is the notion that a primary purpose of penal treatment is to effect changes in the characters, attitudes, and behavior of convicted offenders, so as to strengthen the social defense against unwanted behavior, but also to contribute to the welfare and satisfactions of offenders. (1981, 2)

Rehabilitation, then, relied on a broadly utilitarian justification because punishment on this account produces good results for individual people who have

committed crimes and for society by reducing crime. Based on this idea, novel punishment practices emerged because of a focus on improving the person who committed the crime instead of merely causing them suffering. These new practices included (1) pre-sentencing reports and diagnoses, (2) sentences that were longer and indeterminate because they were defined by treatment goals, (3) probation and parole that allowed for court control even beyond physical custody, and (4) special attention paid to youth who had committed crimes because they were thought to more responsive to rehabilitation than adults (Bailey 2019, 3).[1]

The rehabilitative ideal depended on at least two assumptions. First, it depicted "crime as a social problem that manifested itself in individual acts; individuals became delinquent because they were deprived of education, family socialization, or treatment for their abnormal psychology" (Bailey 2019, 4). Second, it depended on the idea that treatment efforts could change "habits and values of individuals" (4). These assumptions painted a picture of individuals who committed crimes as at the mercy of social circumstances and ultimately dependent on the state's penological responses. While rehabilitation did recognize that crime was connected to poverty and lack of education or social support, rehabilitation sought to treat individuals, not to change the underlying inequalities that produce poverty, lack of education, and unstable social settings. On this medical model, the person who committed a crime was ill, and rehabilitative tools were like medical treatments. Moreover, just as blame was an inappropriate response to illness, blaming those who committed crimes was not appropriate under the rehabilitative model. Instead of evaluating the moral character of a person who committed a crime, the rehabilitative ideal recommended evaluations of mental health, addiction, education, and social training. Punishments came in the form of compulsory mental health treatment, job training, and addiction counseling.

To many readers, this view will sound preferable to the realities of the 2020s, which include mass incarceration, abysmal prison conditions, and heavy stigmatization of those convicted of crimes in the United States (not to mention police brutality). And, in many ways, it was. Professionals in the criminal justice system saw themselves as social workers, and those who were poor or mentally ill were not blamed for the actions that they likely could not have avoided carrying out. But there remained a darker side to this medical model of the cause of crime. Even though they received social services, the individuals convicted of crimes on a rehabilitative model were seen as

[1] While Victor Bailey (2019) describes the rehabilitative ideal in his book, he also critiques the standard story that the rehabilitative ideal was as dominant in the early twentieth century.

defective in some way, be it in terms of mental health, education, or training. While they were not morally blamed, as those convicted of crimes are in the current mainstream view, they were seen as pitiful, in need of rescue by their benevolent betters, and unable to manage their own lives. This was why critics of the rehabilitative ideal argued that it denied the agency of those who committed crimes, treating them as passive participants in their own rehabilitation rather than as responsible individuals.

An even bleaker result of the denial of agency was the concern that some people could not be rehabilitated at all. Indeed, just as rehabilitative sciences were developing, so were programs designed to predict who would continue to commit crimes regardless of treatment. Most criminal sentences were indeterminate, so that one would only be released when they were shown to be recovered. Of course, this meant that many people were simply never released from rehabilitative treatment at all. In cases where rehabilitation was impossible, incapacitation through indefinite incarceration was used as a means of protecting the community (von Hirsch 1985, 5).[2]

2 The Mid-Twentieth Century: Mixed Theories and the Struggle between Medical and Moral Explanations of Crime

Before rehabilitation was completely eclipsed by retribution in the 1970s, philosophers of law were already anxious about the implications of the rehabilitative ideal for concepts of human freedom. These thinkers were still fundamentally concerned with the necessity of punishment for upholding social values like the rule of law and general deterrence, but they were concerned that, without a strict rule in place to ensure that only those who culpably committed crimes be punished, the rehabilitation model would spread past the criminal justice system and become a project of pure social hygiene. Why wait for a person to commit a crime if social scientists could predict that people with certain mental illnesses or social backgrounds were bound to offend—especially if the "punishments" were viewed as treatments designed to cure the person, not to cause suffering? To avoid such implications, so-called "mixed theories" of punishment used utilitarian arguments to justify

[2] Though rehabilitation is not very common penal practice today, it does undergird the civil commitment system in the United States. Indeterminate or permanent incapacitation lives on through this practice, where people who are not convicted due to mental illness or those with some kinds of disorders who are committed without ever committing a crime are incapacitated in "hospitals." See, for example, Hamilton-Smith (2018).

the institution of punishment (usually arguing for the idea that punishment is a necessary deterrent to prevent crime) and retributivist arguments to justify the application of punishment, but only for individuals who are both guilty and morally culpable for committing crimes.[3] The elaboration of Hart's mixed model was a microcosm of the debate between rehabilitation theorists and retribution theorists.

Elaborating the foundations of his mixed model theory about punishment in his Presidential Address to the Aristotelian Society, Hart explained his major concern about theories of punishment that did not take human freedom seriously (2008, 1–27). He divided the justification of punishment into three related questions: (1) "What justifies a general practice of punishment?" (2) "To whom may punishment be applied?" and (3) "How severely may we punish?" (2008, 3). Most theorists, he argued, had tried to answer all three questions with the same theory. But, he argued, one could give a retributivist answer to the second question while still maintaining a utilitarian answer to the first. Importantly, he insisted that there is no utilitarian principle that limited punishment to the guilty, so if one was to be committed to only punishing the guilty, they needed a retributivist theory to answer the second question of punishment. Hart argued that we could accept a guilt requirement in the distribution of punishment, in essence taking the retributivist response to the answer of who gets punished, without being committed to a retributivist answer of why we punish in the first place. Later, the view that guilt was a necessary but not sufficient requirement for punishment would be termed "negative retributivism."

Some utilitarians responded to the criticism that utilitarianism would permit the punishment of innocent people by arguing that, by definition, punishment was the application of suffering in response to the commission of crime, not just any application of suffering. Hart dismissed this argument as a "definitional stop" (2008, 5). The definition stopped us from inquiring into the heart of this criticism: "Why do we prefer [a system of punishing the guilty] to other forms of social hygiene which we might employ to prevent anti-social behavior …?" (6). In other words, if we were going to force some people to be treated for mental illness, addiction, or other "anti-social" tendencies, why wait for them to commit a crime? If the only thing we were after

[3] In the 1950s, John Rawls and H. L. A. Hart both advanced mixed theories of punishment that drew both on utilitarian justifications for punishment and on some retributivist tenets. In 1955, Rawls introduced his approach to punishment in his famous "Two Concepts of Rules." There, Rawls argued that, while the practice of punishment itself was justified by appealing to utilitarian principles, the actual application of the practice through rule to any particular person was justified by the retributive principle of guilt (1955, 4–7).

was reducing crime, as utilitarians suggested, we should have given people rehabilitative treatment as soon as they showed signs that they *might* commit crimes.

Far from being an abstract debate about how definitions work, the definitional stop critique was motivated by highly practical concerns. Hart was worried about proposals by criminologists of his day, led by Lady Barbara Wootton, to use the criminal legal system as a system of social hygiene. Wootton argued that criminal courts could not tell if someone who committed a crime had done so truly voluntarily (Hart 2008, 178–81). She made this argument after a thorough study of how the English courts had made determinations about who had "diminished capacities" and were thus subjected to lesser punishments. She showed that courts either made a circular claim that a person did not have the capacity to follow the law because they regularly committed crimes, or courts simply claimed that having certain categories of mental illness meant that a person had diminished capacity. There was no evidence that directly connected the mental illness with an incapacity to conform oneself to the law. She argued that, while we can make general claims that some kinds of mental illness seemed to occur more often in people who committed crimes, we did not have sufficient evidence that the mental illness was the cause of the crime in any individual case. But, she argued, that did not matter if the best response to all criminal behavior was mental health treatment or other rehabilitative responses. The capacity of the individual was not relevant for determining what kind treatment they needed as "punishment."

Hart found this conclusion troubling because he insisted that we should not punish people who could not have acted differently, even if the punishment itself was rehabilitative. He argued that punishment should track moral blameworthiness as much as possible. In order to be held morally blameworthy for committing a bad act, a person must have had what he called "capacity-responsibility." Hart defined capacity-responsibility thusly:

> "He is responsible for his actions" is used to assert that a person has certain normal capacities. … The capacities in question are those of understanding, reasoning, and control of conduct: the ability to understand what conduct legal rules or morality require, to deliberate and reach decisions concerning these requirements, and to conform to decisions when made. (2008, 227)

A person was only responsible in a moral sense if they could have understood what they were doing (e.g., they were not hallucinating), if they could have decided about a plan of action (e.g., they did not have mental illnesses that impair planning), and if they could have carried out that plan (e.g., they were

not operating under a compulsive disorder). If a person could not have understood a moral rule or was unable to make themselves conform to it, then they could not have been morally responsible. While not every legal system limited criminal liability to those with capacity-responsibility, Hart argued that it was unjust to punish someone without a procedure in place to make sure that they had had capacity-responsibility (2008, 227–30). If Wootton was right that courts could not have actually determined if any particular person who had committed a crime could have acted otherwise, that fact undermined traditional retributivism and Hart's own mixed theory that relied on individual guilt as a necessary condition for legitimate punishment.

Hart insisted that, even in the face of Wootton's evidence that courts did not have the means to determine if a person committed a crime as a result of mental illness, the legal system should not have abandoned the question of whether or not a person could have avoided committing a crime. In the face of uncertainty about whether or not we could have really known if other people were acting with full capacity-responsibility, Hart called for a different approach to thinking about responsibility based on social practices that prioritize human freedom (2008, 181).

For Hart, Wootton's position was not tenable because failing to take the capacity-responsibility of the person who committed a crime into account was at odds with every other aspect of social life. Hart argued that, even if we had some reason to believe that there were times when people, due to mental illness, could not have acted otherwise, we should have assumed, as a general rule, that others did act volitionally:

> Human society is a society of persons; and persons do not view themselves or each other merely as so many bodies moving in ways which are sometimes harmful and have to be prevented or altered. Instead persons interpret each other's movements as manifestations of intention and choices, and these subjective factors are often more important to their social relations than the movements by which they are manifested or their effects. (2008, 182)

No matter how much scientific evidence we may have had that other people's actions were produced by a chemical deficiency, disorder, or other source, Hart argued that we, as human beings, still interpreted the actions of others through the lens of intentionality. It mattered to us, he argued, whether our neighbor accidentally elbowed us in the face or purposefully did so, even if both actions cause a bloody nose.

Thus, Hart insisted that legal systems should have endeavored to determine if a person who committed a crime had capacity-responsibility at the time. In

general, we should have assumed that people did have control over their actions, meaning that they had capacity-responsibility. Only when there was positive evidence that a person did not have capacity-responsibility at the time of the crime should we have reduced their legal liability for that crime.

While Hart argued that Wootton went too far in rejecting the idea that people were morally responsible for the crimes they committed, he bemoaned the reality that most criminal legal systems had not done enough to prevent those with diminished capacity-responsibility from being punished. For example, Hart discussed the M'Naghten Rule, a test to determine if someone was mentally incapacitated in a way that undermined their criminal liability even though they committed a criminal act. This "insanity" test was articulated after Daniel M'Naghten killed the secretary to the prime minister of England in 1843, believing him to be the actual prime minister. M'Naghten believed that the Tories were conspiring to kill him, and his lawyer successfully argued to a jury that he should be found not guilty by reason of insanity. When popular outrage at the not guilty verdict erupted, a new rule, now called the M'Naghten rule, was articulated delineating a relatively narrow criterion for criminal insanity:

> All defendants are presumed to be sane unless they can prove that—at the time of committing the criminal act—the defendant's state of mind caused them to (1) not know what they were doing when they committed said act, or (2) that they knew what they were doing, but did not know that it was wrong. (Legal Information Institute 2022b, "M'Naghten Rule")

This test allowed for a person who had a *cognitive* failing due to mental illness to avoid criminal responsibility. For example, if a person had killed someone while under a delusion that the victim was a violent extraterrestrial bent on destroying humankind, the killer would have been found "not guilty by reason of insanity." However, if a person knew fully well what they were doing but were unable to stop themselves because of a compulsive disorder, they would not have counted as "insane" under the M'Naghten test. This means that a *volitional* failing was not enough to qualify a person as "not guilty by reason of insanity."

During the 1950s and 1960s, some states in the U.S. adopted new tests to remedy this apparent mismatch, most notably the test developed by the American Legal Institute as a part of its project to update and standardize American criminal law through the creation of the Model Penal Code (MPC), the first version of which was published in 1962. Under the MPC test,

> an individual is not liable for criminal offenses if, when he or she committed the crime or crimes, the individual suffered from a mental disease or defect that resulted in the individual lacking the substantial capacity [1] to appreciate the wrongfulness of his or her actions or [2] to conform his or her actions to requirements under the law. (Legal Information Institute 2022a, "Model Penal Code Insanity Defense")

While the first clause was meant to capture the essence of the *cognitive* criterion spelled out in the M'Naghten test, the second clause was meant to offer a *volitional* criterion that would cover instances where a person's mental illness impaired their ability to control their actions but did not impair their ability to comprehend the situation or its normative requirements. The MPC test and other similar tests that included something like the volitional element, such as the Durham test and the irresistible impulse test, were widely adopted in the United States starting in the 1960s. Meanwhile, in the U.K., the Homicide Act of 1957 (5 & 6 Eliz. 2 Ch. 11) did not amend the M'Naghten Rule but instead added a partial defense of diminished responsibility, which mitigated a murder charge to manslaughter.[4] Hart favored the expansion of insanity tests, noting that many European codes included both cognitive and volitional prongs. In contrast, he considered the Homicide Act's provision "both meagre and half-hearted" (2008, 193).

Hart's opponent, Wootton, would have argued that none of these tests could have accurately determined who had capacity-responsibility, and even if they could have, they were not necessary (Hart 2008, 178). But Hart insisted that punishment was only appropriate when a person had both the capacity and a fair opportunity to avoid committing a crime. Although tests for diminished capacity and insanity may have been imperfect, they were still preferable to abandoning the question of capacity-responsibility for crime completely. With this move, Hart was attempting to walk a fine line. On the one hand, he argued that it mattered to us deeply in our daily lives whether or not others acted with intention, which seemed to indicate that we cared if someone could have acted differently when they did something harmful. But social

[4] The Royal Commission on Capital Punishment that convened from 1949 to 1953 officially recommended that the M'Naghten Rule be amended along similar lines as the MPC test. The Commission recommended the following language: "The jury must be satisfied that, at the time of committing the act, the accused, as a result of disease of the mind (or mental deficiency) (a) did not know the nature and quality of the act or (b) did not know that it was wrong or (c) was incapable of preventing himself from committing it" (Royal Commission on Capital Punishment 1953, 111). To count as diminished responsibility under the Homicide Act of 1957, a criminal defendant must show that they were "suffering from such abnormality of mind (whether arising from a condition of arrested or retarded development of mind or any inherent causes responsibility, or induced by disease or injury) as substantially impaired [their] mental responsibility" (Homicide Act of 1957 [5 & 6 Eliz. 2 Ch. 11], Pt. 1, sec. 2).

science research seemed to suggest that we could not know if an act was truly voluntary or not (if any act was every truly voluntary). So, we would have to rely on our social conventions to tell us whether acts were done purposefully or not. We could observe the actions of others for telltale signs that bodily movements were involuntary—a tick, a startle, or other indication of accidental movement—or that the person's actions looked purposeful. We asked people why they did certain actions and evaluated their credibility when they answered. And the criminal law also properly relied on these standards, allowing inferences of capacity based on commonsense reasoning.

Hart wrestled with how to address the question of determining who had a meaningful ability to act otherwise when they committed a crime. His rejection of Wootton's position entailed a rejection of the rehabilitative ideal's model that crime was caused by mental problems or other types of social defects. The only crimes that we were justified in punishing were those that arose out of an individual's agency, so we needed robust tests for legal insanity. But Hart also asserted that, outside the rare times these tests captured, most people had capacity-responsibility most of the time.

3 The 1970s: The Rise of Retributivism and the Moral Description of Crime

While Hart had already raised alarms about rehabilitation's implications for the domain of human freedom in the late 1950s and 1960s, it was not until the 1970s that retributivism replaced the rehabilitative ideal as the dominant theory of punishment. In this "golden half-century" of punishment theory starting in 1957,[5] most punishment theories focused on justifying punishment by appealing to the concept of punishment itself, not on empirical facts about what punishment could achieve (M. Davis 2009). But the myriad of theories that developed under the moniker of "retributivism" were so diverse that the term failed to capture the meaning of all these different theories. Basically, any account that did not rely primarily on empirically measurable social outcomes tended to fall under this label.[6] I cannot give an account of each of these wide varieties of retributivism here, so I will focus on three theorists that illustrate the retributivist assumption that crime is primarily a moral

[5] Michael Davis (2009) traced the beginning of the rise of retributivism all the way back to 1957, while von Hirsch (1985, 9) argued that 1971 was when serious retributivist theory took root.

[6] Michael Davis helpfully argued that we should shed the utilitarian/retributivist distinction and replace it with an empirical/conceptual conception (2009, 89).

problem. These three theories highlight important shared values with the "law and order" movement that many scholars of incarceration link to mass incarceration.

While there were likely many causes of its decline, empirical evidence that the penal practices meant to rehabilitate did not in fact produce the intended results of reducing recidivism fueled many critiques of rehabilitative models (Alschuler 2003, 9). Andrew von Hirsch summarized: "The results were disappointing, indeed. Although many offenders seemed to show improvement (that is, did not return to crime), this tended to occur as much among untreated as among treated individuals—the treatment as such had little perceptible influence" (1985, 4).[7] Moreover, between 1960 and 1980, crime rates rose suddenly and quickly. Property crime rates rose by 200 percent and violent crime rates rose by about 250 percent (Pfaff 2017, 3).

Deterrence, a different utilitarian ideal of punishment, briefly took center stage for punishment theorists. In the 1970s, as rehabilitation theories declined, von Hirsch argued that general deterrence took on an outsized role in punishment, particularly as philosophers and jurists appealed to law and economics modeling (von Hirsch 1985, 7–9). While still aiming to reduce crime overall, this approach instead viewed those who might commit crimes as rational agents who would take harsher penalties into account. Just as with rehabilitation, the role of incapacitation paired nicely with the aim of general deterrence, as some criminologists thought that much violent crime was perpetrated by a handful of repeat offenders. Thus, long prison sentences had the double effect of taking repeat offenders off the streets and sending a strong deterrent message to those members of society who would respond to rational incentives (von Hirsch 1985, 7–9).

According to von Hirsch, deterrence fit the new "law and order" attitude that was emerging in the 1970s in the United States (von Hirsch 1985, 9). Indeed, the 1970s marked the point when the incarceration rate in the U.S. broke from historically stable rates, slowly increasing as the decade went on. At the beginning of the 1970s, the rate of incarceration was in line with what it had been since the mid-1800s and broadly consistent with the incarceration rates of the U.K. and European countries (Pfaff 2017, 1). During the 1970s, scholars tended to think that the incarceration rate would either radically decrease from its already low rate because of continuous prison reformation projects (Rothman 1990, 295) or remain steady indefinitely because states would adjust policies to maintain relatively consistent incarceration

[7] For more information on the decline of the rehabilitative ideal, see Secherest et al. (1979), Allen (1981), and Bailey (2019).

rates (Blumstein and Moitra 1979, 376). But both would soon be proven wrong. To highlight the contrast, at a low water mark, in 1972, fewer than 200,000 people were incarcerated in state and federal facilities, but that number jumped to 1.56 million in 2014, not including county jails (Pfaff 2017, 2).

It is impossible to draw causal connections one way or the other between philosophers and policymakers, but there are certainly affinities between some theories and the public policies that led to mass incarceration. Von Hirsch may have been correct to note that the deterrence theorists he highlighted seemed to align with the growing "law and order" mentality in the United States. But retributivist models shared a distinctive emphasis on individual moral responsibility with the "tough-on-crime" political rhetoric of the 1980s and 1990s, the same decades that saw huge increases in incarceration rates. By that point, retributivist theories, particularly those focusing on moral desert, had established predominance in philosophy of punishment.

At first, it seemed that retributivist theories were actually meant to be kinder, more balanced approaches to punishment. In contrast to deterrence, desert-based theories of punishment held that the central purpose and justification of punishment was to give those who had committed crimes what they deserve. The severity of punishment (particularly the length of a prison sentence) was meant to match the severity of the crime committed. Sentences were not to be designed to rehabilitate the person who committed the crime or deter others from committing the same crime. In the face of sentences that were indeterminate (rehabilitative) or excessively long (deterrence-based), von Hirsch argued that the desert-based models were introduced as justice-centric interventions to limit extreme or indeterminate sentences (von Hirsch 1985, 11).

Moreover, as Hart's arguments against Wootton emphasized, retributivism entailed a deep respect for the agency of individual human beings. While there may be real cases in which mental illness or other problems prevented people from acting freely, retributivists were wary of granting too many exceptions to the rule that people were agents who were responsible and blameworthy for criminal actions. For many retributivists, moral desert and strong blaming practices went hand in hand.

But, just as rehabilitation had a dark side, moralistic retributivism also had pernicious views of individuals who commit crimes. One of the most moralistic versions of retributivism that arose in response to the rehabilitative ideal was Michael S. Moore's theory. In an essay first published in 1987, he explained, "Retributivism is a very straightforward theory of punishment: We are justified in punishing because and only because offenders deserve it" (2010, 181). For Moore, guilt was sufficient on its own sufficient to justify

punishment. To defend the idea that guilt was sufficient for warranted punishment, Moore argued that we should pay attention to our intuitions when we hear about the commission of heinous crimes. To motivate this argument, he quoted at length from a 1981 editorial by Mike Royko, a long-time columnist for the *Chicago Times*.[8] In the column, Royko explained his outrage at anti-death penalty advocates who sang "We Shall Overcome" outside a prison that held murderer Steven Judy. Royko could not sympathize with the protestors because he had met living relatives of a number of murder victims, including Judy's victim. Royko detailed many other violent and shocking crimes alongside empathetic portrayals of victims and their families (Royko 1981).

This newspaper column was the start of Moore's argument. Moore continued, arguing that most people would have the immediate intuition that the people who committed the crimes that Royko describes should be punished (and harshly). That immediate intuition was an important source of knowledge about what punishment was deserved. But, Moore said, most people would then correct their initial response by adding that the reason to punish was to deter, rehabilitate, or incapacitate. This invalidation of the retributivist intuition was where people went wrong. Moore asserted that the emotions that attend retributivism were not misleading. These emotions gave us helpful information to form our moral judgments.

In another appeal to intuitions, Moore again turned to an example of a real, heinous crime to motivate his argument. This time, he asked readers about their intuitions about a jilted boyfriend, Herrin, who murdered his girlfriend, Garland, with a hammer (2010, 213). Moore argued that we should all agree that we could see ourselves ending up making a horrible mistake with brutal consequences, just like Herrin. A failure to understand this truth could only

[8] Royko's disdain for criminals also took on racial tones in his 1993 article contrasting Rodney King to Barbara Meller Jensen, a German woman who was murdered when she got lost and ended up in a "low-income, high-crime area" (Royko 1993). He went on to complain that King's beating got too much attention, while Jensen's murder was quickly forgotten:

> Despite his troubles, Rodney King is a lucky guy. He is a criminal by trade, having served time for armed robbery. He was beaten after driving drunk and being chased at dangerously high speeds, putting innocent motorists at risk. He was a social menace. In contrast, Mrs. Jensen was a law-abiding, useful person: a therapist for handicapped children. She had gone to Florida because her husband, a biologist, needed solitude to complete a book. She wasn't a threat to anyone. (Royko 1993)

Of course, one cannot impute every view that Royko articulated in his decades as a well-loved columnist to Moore just because Moore quoted him at length in one paper in 1988. But the sentiment of these two columns is consistent. Those who commit crimes deserve harsh punishments, not our sympathy. Because we care about the victims of crimes, we are justified in retributive feelings, and the institutions of the state should carry out punishments in accordance with these retributive feelings.

be attributed to those people who have made a deep "we-they" distinction between themselves and those who commit crimes. Knowing that anyone was capable of violence in the right circumstances, he followed up:

> Then ask yourself: What would you feel like if it was you who had intentionally smashed open the skull of a 23-year-old woman with a claw hammer while she was asleep, a woman whose fatal defect was a desire to free herself from your too clinging embrace? My own response, I hope, would be that I would feel guilty unto death. I couldn't imagine any suffering that could be imposed upon me that would be unfair because it exceeded what I deserved. (Moore 2010, 213)

At first, he invited the reader to feel the pain of having killed another person. Intuition told readers what murderers deserve, which they could trust because it was what they would ask for if they committed such a crime.

Next, Moore described Herrin's interview with a psychiatrist in which Herrin asserted that his eight-year sentence was too long, and that he should have been let out after two years, considering his personal circumstances, including no history of prior crimes. To those who would have argued that guilt was an unhelpful or even destructive emotion, Moore responded that feeling guilty and wanting punishment were a much better alternative than Herrin's "shallow, easily obtained self-absolution" (2010, 214). Moore was not arguing for a rehabilitative aim. He did not think Herrin should be punished in order to induce a proper sense of guilt. Instead, Moore deplored the lack of blame and personal responsibility that Herrin's comments exhibited. The fact that we would all presumably feel that no punishment was too great for us if we were in Herrin's shoes was itself reliable evidence for us make judgments about what appropriate punishment was for others.

Moore insisted on the value of individual responsibility, rejecting Herrin's view that his background circumstances made him less culpable. A refusal to consider factors that might undermine the capacity of an individual to make better choices was paradigmatic of the "law and order" era. In a reversal of the 1960s legal movement to broaden the category of "insanity," in the 1980s, there was a massive movement in the United States to return to the more restrictive M'Naghten test. Again, an assassination attempt motivated the narrowing of the insanity defense. John Hinkley, who had attempted to assassinate President Ronald Reagan, was found "not guilty by reason of insanity" in federal court, which used the broader MPC test. Following public outcry, U.S. Congress responded to Hinkley's verdict by changing federal law, returning to the narrower M'Naghten formulation of "insanity." Most states quickly did the same, and five states even passed legislation to abolish the insanity

defense altogether (Morse 2021, 2–3). The narrowing of excuses for crimes was an essential feature of both of Moore's retributivism and the 1980s "law and order" movement.

This attitude was also seen in the rhetoric of politicians. While Republican leaders had often been associated with the tough-on-crime attitudes and rhetoric, many Democratic leaders in the 1990s expressed the same sentiments. Take current President Joe Biden as an example. When Biden was a senator in the 1990s, he was known for being tough on crime. When proposing stringent new crime bills, he talked about "predators on our streets" in a speech in 1993. In 1994, he lauded Nixon's criminal justice policies: "Every time Richard Nixon, when he was running in 1972, would say, 'Law and order,' the Democratic match or response was, 'Law and order with justice'—whatever that meant. And I would say, 'Lock the S.O.B.s up.'" And like Moore, he had no patience for hearing about the social backgrounds of those who commit crimes: "It doesn't matter whether or not they're the victims of society. I don't want to ask, 'What made them do this?' They must be taken off the street," Biden said in 1993 (Stolberg and Herndon 2019).

While Moore's retributivism was especially stark in its emphasis on heinous crimes, strong moral condemnation of any person who committed a crime, and limiting excuses, other retributivists drew similar conclusions by emphasizing the moral nature of criminal wrongs. Jean Hampton (1992) used an expressive retributivism that also encompassed a moral education view to argue that criminal punishments were the best moral responses to sexist and racist crimes. Both her retributivism and her moral education justifications came out of what she saw as the expressive capacities of criminal acts and punishments. She argued that when a person committed a crime, they expressed the idea that they were more important that victims or society at large (1666). The act itself communicated this message by diminishing the victim. This was true of the most heinous crimes. Hampton used the example of a particularly violent anti-Black hate crime in which a White farmer tortured, killed, and mutilated five Black farm hands, a man and his four sons, in response to some minor slight (1675). But it also can be true of the most minor infraction. Hampton used the example of a person who snuck a book out of the university library without checking it out (1680). This person's act announced that they thought their own unfettered access to the book was more important than that of the other members of the university.

For Hampton, punishment was necessary to right these diminishing wrongs. It sent a message that the person who committed the crime was not above the law, and that they were not more valuable than their victim or the community. Because of its capacity to communicate this moral message,

Hampton also thought that punishment could morally educate the person who had committed the crime. But Hampton insisted on a sharp distinction between rehabilitation and moral education: "Apart from any literacy problems, occupational problems, or mental problems, this view [moral education] holds that those who are guilty of a criminal offense have a moral problem" (1998, 40). Punishment sent a moral message that the convicted person could chose to accept or not. It did not, as with rehabilitation, act on a passive patient who had no real agency in their own crimes or their potential redemption.

While Hampton often expressed skepticism about the role of incarceration in punishment in North America, she also argued in favor of limiting a criminal's right to vote based on her emphasis on crime's moral wrongs to victims. She argued that allowing White men who had committed anti-Black hate crimes to vote would communicate the idea that these men are still members of the political community, even if they expressly denied a political value of the equality of all members of the community regardless of race. Likewise, Hampton argued that those who committed violence against women expressed the view that women are less valuable members of society:

> To hand the levers of political power over to someone whose behavior manifests an intention to accomplish the subordination of women to men undermines not only the democratic value of equality but also the status and safety of women in that society. (1998, 42)

Because crimes sent messages about the value of their victims, failing to punish crimes sent a message that the government (and the political community it represented) did not care about women's well-being either.

Hampton's view of criminal punishment as a tool to fight women's oppression was also consistent with the spirit of the tough-on-crime 1990s. Beth R. Richie argued that the women's anti-violence movement split in the 1980s, as some feminists pragmatically chose to align themselves with the "law and order" movement (2012, 84–86). Framing violence against women as a criminal justice issue (rather than a more widespread social issue), these groups secured resources and had a large impact on legislation. The most well-known and significant piece of this legislation was the Violence Against Women Act, first passed in 1994, which foremost provided criminal justice tools for fighting violence against women, and also added funding for shelters and other victims' services. It is less well-known, however, that it was passed as part of the now infamous Omnibus Crime Bill (Pub.L. 103–322), which, among other things, created sixty new federal capital crimes, ended Pell Grants for

prisoners, and provided extensive funding for new prisons and police programs (Richie 2012, 84–86). Hampton's feminist, expressivist retributivism aligned with some feminists' attempts to use the criminal law's power to condemn as a tool to fight gender-based violence. As Richie noted, this strategy was ultimately only successful for middle class, White, heterosexual women (2012, 1–4).

Another moralistic retributivist, Herbert Morris (1968), argued for the right to be punished as a recognition of one's capacity as an agent, offering an argument similar to Moore's. Like Moore, Morris imagined that if he were the one who had committed a heinous crime, he would accept any punishment, so we should want that for others. To make this point, he argued that, if we saw ourselves as potential or actual criminal wrongdoers, we would want a system of punishment that would allow us redemption. Speaking of the person who committed a crime, Morris wrote, "Further, the evil … that he has done himself by his wrongdoing is a moral evil greater than he has done others. His soul is in jeopardy as his victim's is not" (1981, 267). In a softer tone than Moore, Morris argued that, "but for the grace of God," we could all commit crimes that we ourselves abhor. We would have all wanted a path to redemption and reconciliation, so we should have wanted that for others. Punishment offered this path.

Morris proposed a softer version of retributivism than Moore, but his conclusion was the same. Individuals who committed crimes had deep moral problems that only punishment could address. We should have punished these individuals regardless of whether that punishment reduced crime or helped anyone. Of course, we cannot draw a causal connection between retributivist theorists such as Morris, Hampton, or even Moore and mass incarceration or the increasingly obvious injustices of the American criminal legal system. But they shared a certain spirit: the idea of moral responsibility for crime, the blameworthiness of the person who committed it, and the dismissal of factors that might mitigate culpability. Retributivism was the dominant theory of punishment in the 1980s and 1990s, when incarceration saw its sharpest increases, new tough-on-crime legislation passed, and legal defenses shrunk.

In contrast to the more explicitly moralistic desert-based theories, another type of retributivist theory, the "benefits and burdens" theory (also called the "fair play" theory) situated criminal punishment in the context of a larger social-political system.[9] Roughly speaking, those who endorsed this view recognized that society was a cooperative endeavor and that laws were required

[9] Other benefits and burdens theories included Morris (1968), Murphy (1973), and Sher (1987).

to maintain this cooperation. Because members got the benefits of the laws—that is, society ran smoothly enough for them to live comfortably within it—each member was required to take their fair share of the burdens. Those burdens included following the law. When a person broke the law, they took unfair advantage of this balance, and punishment was necessary to reset the just distribution. This may have had the ring of deterrence to it, but for benefits and burdens theorists, resetting the balance was a moral response to crime, not an instrumental one.

On Richard Dagger's (1993) version of a benefits and burdens argument, a system of punishment secured social cooperation by deterring would-be rule-breakers and assuring those who want to follow the rules that they will not be taken advantage of. This system of benefits and burdens created obligations for each person to follow the law:

> Criminals act unfairly when they take advantage of the opportunities the legal order affords them without contributing to the preservation of that order. In doing so, they upset the balance between benefits and burdens at the heart of the notion of justice. (Dagger 1993, 476)

Thus, each crime committed was morally wrong not just if the act itself was morally wrong outside the law (murder, assault), but it was also wrong because it violated a moral obligation to support the system of social cooperation. Moreover, this meant that in committing a crime, one did not just break a moral duty to a specific victim, if there was one, but also to every member of society.

One interesting upshot of the benefits and burdens version of retributivism was that deep injustice in the system of social cooperation undermined the justification of punishment: "And this means that punishment is justified only when there is a just balance of benefits and burdens to begin with—when the social order is just, or reasonably so" (Dagger 1993, 177). We could not blame a person for upsetting a fairly balanced set of benefits and burdens if there was not a fair equilibrium to start with. With this in mind, Jeffrey Reiman argued, "Since the obligation to obey the law is a function of the benefits one receives, it follows that many disadvantaged criminals are not violating their moral obligations to obey the law" (2007, 7–8). Thus, in the United States and other similar political communities, many people who have committed crimes were not morally obligated to follow the law itself and did not merit state-based punishment (although they may still have been morally obligated, to, say, avoid assault or murder).

This version of retributivism was a vast improvement on desert-based theories in light of the concerns that continue to plague the American criminal legal system in the 2020s. Benefits and burdens theories left room for thinking about how background injustices affected the obligations that members of a society had toward one another. They raised questions about the legitimacy of punishing oppressed groups, including those living in poor Black neighborhoods that were especially impacted by crime, violent policing, and incarceration.[10] But, at the same time, they shared much with retributivism's more classic version. Even if a person should not have been punished because of background injustice, the framework was still about individual obligations to follow the law and whether those obligations applied.

4 Concluding Remarks on the Social Critical View of Crime

Throughout the twentieth and twenty-first centuries in the United States, social critics of punishment have led social movements, sometimes within prisons themselves (Adelsberg et al. 2015). Angela Davis, a trained philosopher and social movement leader, has long been a part of anti-prison movements. In her abolitionist text, she asks "why 'criminals' have been constituted as a class, and indeed a class of human beings undeserving of the civil and human rights accorded to others" (2003, 112). The social critical view of crime is that it is caused by deep injustices in the basic structure of society, including racial and gender oppression, poverty and inequality, and ablism. Viewing crime as primarily a result of structural injustice means that dealing with crime through individual punishments cannot address the problem of crime either practically or morally. Criminal law functions to maintain group-based oppressions, regardless of the intentions of individuals who carry out the tasks of the criminal justice system. In the United States, Black Americans, especially those descended from enslaved people, were (and continue to be) particularly targeted by the criminal law (Alexander 2012). Moreover, the vast majority of people of all races who end up being incarcerated were (and continue to be) extremely poor.

Whereas the Black Lives Matter protests starting in 2014 often included calls for particular officers to be arrested and convicted, in 2020, many Black Lives Matter protesters started to demand that police departments be

[10] For a nuanced discussion of the legitimacy of punishing poor Black people in the United States, see Shelby (2016).

defunded, appealing to similar themes from prison abolitionists like Davis. Protestors called for a shift in funding from police to education, healthcare, housing, and other public services. At first glance, the contemporary calls for ending policing and incarceration are similar to the old rehabilitative model. Like the rehabilitative model, the social critical view sees social and political problems rather than moral problems as the root of crime. It also emphasizes the value of social services and recognizes that those who live in poverty have many more reasons to commit crimes than those with financial stability.

But the social critical view is fundamentally different than the rehabilitative model because it does not focus on compelling those who have committed crimes to accept services such as mental health care, job training, and addiction treatment. Rather than addressing crime by treating those who committed crimes as patients receiving care from benevolent experts, the social critical model emphasizes the active involvement of communities that have been excluded from political life and most impacted by crime and punishment. The state is complicit in the harms that crime causes because it contributes to income inequality, racial disparities, and lack of access to health care and income. In addition to increasing mental health care and substance abuse treatment access, as was typical of rehabilitation, social critics call for addressing crime at its root. This means fighting poverty though measures such as stronger progressive taxation, welfare, or universal basic income; providing free higher education and job training; and investing in infrastructure in Black, Indigenous, and Latinx communities. As social movements and policy makers move into the next decade of addressing crime and punishment, the social critical view of crime is a powerful philosophical approach to thinking about punishment and justice. Present-day activists and theorists who justly demand the return of vital social services, which were dismantled with the rehabilitative ideal, should do so while being wary of repeating the mistakes of the rehabilitative model. But we should not understand the social critical view or abolition and defunding movements as merely demanding more social services.

References

Adelsberg, Geoffrey, Lisa Guenther, and Scott Zeman. 2015. *Death and Other Penalties: Philosophy in a Time of Mass Incarceration*. New York: Fordham University Press.

Alexander, Michelle. 2012. *The New Jim Crow: Mass Incarceration in the Age of Colorblindness*. Rev. ed. New York: New Press.

Allen, Francis A. 1981. *The Decline of the Rehabilitative Ideal: Penal Policy and Social Purpose*. New Haven, CT: Yale University Press.

Alschuler, Albert W. 2003. "The Changing Purposes of Criminal Punishment: A Retrospective on the Past Century and Some Thoughts about the Next." *University of Chicago Law Review* 70, no. 1 (Winter): 1–22.

Bailey, Victor. 2019. *The Rise and Fall of the Rehabilitative Ideal, 1895–1970*. Abingdon, UK: Routledge.

Blumstein, Alfred, and Soumyo Moitra. 1979. "An Analysis of the Time Series of the Imprisonment Rate in the States of the United States: A Further Test of the Stability of Punishment Hypothesis." *Journal of Criminal Law & Criminology* 70, no. 3 (Autumn): 376–90.

Dagger, Richard. 1993. "Playing Fair with Punishment." *Ethics* 103, no. 3 (April): 473–88.

Davis, Angela Y. 2003. *Are Prisons Obsolete?* New York: Seven Stories.

Davis, Michael. 1990. "Recent Work in Punishment Theory." *Public Affairs Quarterly* 4, no. 3 (July): 217–32.

———. 2009. "Punishment Theory's Golden Half Century: A Survey of Developments from (about) 1957 to 2007." *Journal of Ethics* 13, no. 1 (January): 73–100.

Feinberg, Joel. 2008. "The Classic Debate." In *Philosophy of Law*, 8th ed., edited by Joel Feinberg and Jules Colman, 625–29. Belmont, CA: Wadsworth.

Hamilton-Smith, Guy. 2018. "The Endless Punishment of Civil Commitment." *Appeal*, September 4. https://theappeal.org/the-endless-punishment-of-civil-commitment/.

Hampton, Jean. 1992. "Correcting Harms versus Righting Wrongs: The Goal of Retribution." *UCLA Law Review* 39, no. 6 (August): 1659–702.

———. 1998. "Punishment, Feminism, and Political Identity: A Case Study in the Expressive Meaning of the Law." *Canadian Journal of Law and Jurisprudence* 11, no. 1 (January): 23–45.

Hart, H. L. A. 2008. *Punishment and Responsibility: Essays in the Philosophy of Law*. 2nd ed. Oxford: Oxford University Press.

Legal Information Institute. 2022a. "Model Penal Code Insanity Defense." *Wex Legal Dictionary*. Cornell Law School. Accessed April 6. https://www.law.cornell.edu/wex/model_penal_code_insanity_defense.

———. 2022b. "M'Naghten Rule." *Wex Legal Dictionary*. Cornell Law School. Accessed April 6. https://www.law.cornell.edu/wex/m%27naghten_rule.

Moore, Michael S. 2010. "The Moral Worth of Retribution." In *Responsibility, Character, and the Emotions: New Essays in Moral Psychology*, edited by Ferdinand Schoeman, 179–219. Cambridge: Cambridge University Press.

Morris, Herbert. 1968. "Persons and Punishment." *Monist* 52, no. 4 (October): 475–501.

———. 1981. "A Paternalistic Theory of Punishment." *American Philosophical Quarterly* 18, no. 4 (October): 263–71.

Morse, Stephen J. 2021. "Before and after Hinckley: Legal Insanity in the United States." *Faculty Scholarship at Penn Law*. 2252. https://scholarship.law.upenn.edu/faculty_scholarship/2252.

Murphy, Jeffrie G. 1973. "Marxism and Retribution." *Philosophy & Public Affairs* 2, no. 3 (Spring): 217–43.

Pfaff, John F. 2017. *Locked In: The True Causes of Mass Incarceration—and How to Achieve Real Reform*. New York: Basic.

Rawls, John. 1955. "Two Concepts of Rules." *Philosophical Review* 64, no. 1 (January): 3–32.

Reiman, Jeffrey. 2007. "The Moral Ambivalence of Crime in an Unjust Society." *Criminal Justice Ethics* 26, no. 2 (Summer/Fall): 3–15.

Richie, Beth E. 2012. *Arrested Justice*. *Arrested Justice: Black Women, Violence, and America's Prison Nation*. New York: New York University Press.

Royko, Mike. 1981. "Nothing Gained by Killing a Killer? Oh Yes, There Is." *Los Angeles Times*, March 13.

———. 1993. "Rodney King and a Double Standard." *Chicago Daily Tribune*, April 6. https://www.chicagotribune.com/news/ct-xpm-1993-04-07-9304070056-story.html.

Rothman, David J. 1990. *The Discovery of the Asylum: Social Order and Disorder in the New Republic*. Rev. ed. Boston: Little, Brown.

Royal Commission on Capital Punishment. 1953. *Royal Commission on Capital Punishment, 1949–1953: Report*. London: H. M. Stationery Office.

Sechrest, Lee, Susan O. White, and Elizabeth D. Brown, eds. 1979. *The Rehabilitation of Criminal Offenders: Problems and Prospects*. Washington, DC: National Academy of Sciences.

Shelby, Tommie. 2016. "Punishment." In *Dark Ghettos: Injustice, Dissent, and Reform*, 228–51. Cambridge, MA: Belknap.

Sher, George. 1987. *Desert*. Princeton, NJ: Princeton University Press.

Stolberg, Sheryl Gay, and Astead W. Herndon. 2019. "Lock the S.O.B.'s Up: Joe Biden and the Era of Mass Incarceration." *New York Times*, June 25. https://www.nytimes.com/2019/06/25/us/joe-biden-crime-laws.html.

von Hirsch, Andrew. 1985. *Past or Future Crimes: Deservedness and Dangerousness in the Sentencing of Criminals*. New Brunswick, NJ: Rutgers University Press.