The Dead End of Deterrence, and Beyond

Kyron Huigens
khuigens@yu.edu
THE DEAD END OF DETERRENCE, AND BEYOND

KYRON HUIGENS*

INTRODUCTION

The deterrence theory of punishment has been taken a long way down a new path. The descendants of Bentham now argue that the criminal law addresses "the moral goodness of citizens' motivations." The heirs of Hume and Holmes maintain that the criminal law is a "[p]reference-[s]haping [p]olicy." These developments are as welcome as they are surprising. One wonders, however, if the deterrence theorists who have taken the new path know where they are, how they got there, or where they are going. Dan Kahan, the leader of this hardy band of scholars, eschews "deep theorizing," and scoffs at the notion that some "abstract 'theory' supplies the 'truth' about criminal law." As it happens, Kahan and company inadvertently have taken deterrence theory to its inevitable dead end. It is time to move beyond the notion of deterrence to a new conception of punishment's justifying purpose.

* Associate Professor of Law, Benjamin N. Cardozo School of Law. The author thanks George Fletcher, Jeanne Schroeder, Scott Shapiro, and Stewart Sterk for their helpful comments on earlier drafts. Thanks also are due to the members of the William & Mary Law Review for their editorial assistance.

5. Kahan, supra note 1, at 2496.
This Article consists of a brief statement of a virtue ethics theory of punishment, which follows an extended examination of three versions of deterrence theory: those of Paul Robinson, H.L.A. Hart, and Kahan. The logic of this organization can be stated briefly. The virtue ethics theory of punishment answers a question—What is the nature and significance of fault in just punishment?—to which deterrence theory never has offered a persuasive answer. Kahan's "new path" deterrence theory almost gets it right. This Article picks up where Kahan's work leaves off.

The dominant consequentialism of Anglo-American legal theory gives the deterrence theory of punishment credibility, but this

6. Whereas the retributive theory of punishment has its roots in Kantian deontological ethics, and deterrence theory has its roots in consequentialism, the third major tradition in philosophical ethics—Aristotelian virtue ethics—has not been widely recognized as a viable basis for criminal law theory. The view prevails that "[t]he dominant approaches to justification [of punishment] are retributive and utilitarian." PAUL H. ROBINSON, FUNDAMENTALS OF CRIMINAL LAW 32 (2d ed. 1995) (quoting Kent Greenawalt, Punishment, in 4 ENCYCLOPEDIA OF CRIME AND JUSTICE 1336, 1337 (Sanford H. Kadish ed., 1983)); see also JOSHUA DRESSLER, CASES AND MATERIALS ON CRIMINAL LAW 21-35 (1994) (presenting only utilitarian and retributive justifications for punishment by means of selections from, inter alia, Bentham and Kant, respectively); SANFORD H. KADISH & STEPHEN J. SCHULHOFER, CRIMINAL LAW AND ITS PROCESSES: CASES AND MATERIALS 102-19 (6th ed. 1995) (same); RICHARD G. SINGER & MARTIN R. GARDNER, CRIMES AND PUNISHMENT: CASES, MATERIALS, AND READINGS IN CRIMINAL LAW 84-124 (2d ed. 1996) (same).

Nevertheless, there have been scattered efforts in the direction of a virtue ethics theory of punishment. See George P. Fletcher, The Fall and Rise of Criminal Theory, 1 BUFF. CRIM. L. REV. 275, 287 (1998) ("The utilitarians and Kantians have in fact had much to say about the rationale for punishment. Virtue theorists have recently offered us a more subtle account of culpability.") (citing Kyron Huigens, Virtue and Inculpation, 108 HARV. L. REV. 1423 (1995), and Dan M. Kahan & Martha C. Nussbaum, Two Conceptions of Emotion in Criminal Law, 96 COLUM. L. REV. 269 (1996)); see also Peter Arenella, Character, Choice and Moral Agency: The Relevancy of Character to Our Moral Culpability Judgments, in CRIME, CULPABILITY, AND REMEDY 59, 60-61 (Ellen Frankel Paul et al. eds., 1990) (arguing that so-called rational choice theory is inadequate to describe the criminal law's concern with character); Edmund L. Pincoffs, Legal Responsibility and Moral Character, 19 WAYNE L. REV. 805, 918 (1973) (arguing that punishment represents a demand that one develop and exhibit certain character traits).

7. The term "deterrence theory of punishment" is a misnomer, as is the term "retributive theory of punishment." I will use these standard terms in this Article, but it should be noted that the terms "consequentialist theory of punishment," and "deontological theory of punishment," respectively, are more accurate. Both deterrence and retribution are merely functions of punishment, as are incapacitation, rehabilita-
theory has always suffered from a prominent weakness. Fault—also known as desert, culpability, or blameworthiness—is the distinctive feature of the criminal law, but consequentialism has no independently viable conception of fault. Fault is an irreducibly retrospective concept, and the inveterately prospective orientation of deterrence theory’s underlying consequentialism cripples its efforts to give an adequate account of fault.

One possible response to this evident shortcoming in deterrence theory is simply to adopt and adapt the retributivist conception of fault as intentional wrongdoing—a strategy that Paul Robinson has developed. The requirement of intentional wrongdoing can be imposed as a side-constraint on punishment, for reasons that are distinct from punishment’s justification, but

8. My reasons for using “fault” instead of the more commonly used terms are explained below. See infra note 38.


10. See Alan Brudner, Agency and Welfare in the Penal Law, in ACTION AND VALUE IN THE CRIMINAL LAW 21, 43 (Stephen Shute et al. eds., 1993) (“Accordingly, the doctrine of mens rea has no role to play in the theoretical account of punishment within the welfarist paradigm; it is not part of an account of the inner necessity and deservedness of punishment, because there is here no inner necessity or deservedness to comprehend.”); see also Jules L. Coleman, Crime, Kickers, and Transaction Structures, in CRIMINAL JUSTICE 313, 323-26 (J. Roland Pennock & John W. Chapman eds., 1985) (arguing that the concept of fault simply is absent from economics); Louis Michael Seidman, Soldiers, Martyrs, and Criminals: Utilitarian Theory and the Problem of Crime Control, 94 YALE L.J. 315, 320 n.11 (1984) (noting that the economic analysis of crime suggests “no inherent reason why crimes themselves should be defined with respect to any mental element”).

11. See infra notes 106-19 and accompanying text.
that are perfectly sound. However, this side-constraint interpretation of fault and the liberal and efficiency-minded reasons given in support of it cannot explain the frequent and apparently just punishment of nonintentional wrongdoing that is not merely strict liability. The explanation of nonintentional fault requires an understanding of genuine fault—"genuine" in the sense that it is an affirmative, justifying reason to punish in a given case, and not only a necessary condition. This conception of fault, like any adequate conception of fault, lies beyond the theoretical resources of deterrence theory.

H.L.A. Hart takes a second, fundamentally different, and admirably straightforward approach to the question of fault in the criminal law. He denies forthrightly that fault in the sense of "moral culpability" is either a necessary condition for punishment or an affirmative, justifying reason to punish. Hart retains the requirements of voluntariness and the absence of excusing conditions, such as duress and insanity, as side-constraints on punishment. He calls these the "criteria of responsibility," and he collapses mens rea elements of proof into the same liability-limiting category. One result of these steps is a persuasive defense of nonintentional criminal liability in the form of negligence liability. But even in his deservedly influential analysis of negligence, Hart deliberately denies the existence of nonintentional fault—indeed, of any kind of fault. I rely on J.L. Austin's classic essay, *A Plea for Excuses*, to argue that this rejection of fault is untenable.

Hart might have developed, rather than obscured, the concept of fault. Hart argued elsewhere that law does not govern only

---

12. See infra notes 120-52 and accompanying text.
16. See HART, supra note 14, at 31-37.
18. See infra notes 169-80 and accompanying text.
by way of conscious instrumental reasoning, and that people can be governed by legal rules at the level of their motivations. ¹⁹ This idea can premise the theoretical account of genuine non-intentional fault that deterrence theorists have failed to provide. Kahan's "new path" deterrence theory is important because it picks up this promising line of thought, albeit not from Hart. Kahan attempts to describe the ways in which the criminal law shapes motivations, and in his account of mistake he has sketched the outlines of a deterrence theory of genuine fault. ²⁰ Kahan's efforts in this direction fall short. ²¹ He, too, fails to develop the concepts of genuine fault and nonintentional fault, because he stays within the confines of consequentialist ethics. Under the less than benign influence of the law and economics movement, he fails to work out the full implications of his own insights concerning value, practical reasoning, and the criminal law. Kahan's difficulty is that he has adopted Lawrence Lessig's theory of the regulation of social meaning as a basic theoretical structure. ²² Lessig's theory itself is an attempt to cross the economic analysis of law with certain conceptions of value and practical reasoning that, heretofore, have had no place in economic thinking. ²³ Lessig's difficulty is that these notions come from the nonconsequentialist tradition in philosophical ethics, and they simply cannot be incorporated into economic analysis, which is an inveterately consequentialist enterprise. The incoherence of Lessig's theory obstructs Kahan's "new path" deterrence theory.

Yet, the difficulties of Kahan and Lessig are not my main concern. I am interested in the nonconsequentialism that they fail to use persuasively. The features of nonconsequential value and practical reasoning that interest Kahan and Lessig form the foundation of virtue ethics. ²⁴ Virtue ethics is an ancient body of

¹⁹. See Hart, supra note 14, at 44.
²⁰. See Kahan, supra note 2, at 136-52.
²¹. See infra notes 181-291 and accompanying text.
²⁴. See infra notes 292-307 and accompanying text.
thought that begins with Aristotle's *Nicomachean Ethics*, but that has been revived in philosophical ethics only within the last fifty years. It has never played a significant role in modern legal theory, which has been dominated for at least the last one hundred years by consequentialism. This, however, is what makes Kahan's and Lessig's efforts so interesting and, for my purposes, so useful. Their efforts at hybridization demonstrate the inadequacy of consequentialist legal analysis and of law and economics in particular; and demonstrate, at the same time, the plausibility and importance of nonconsequentialist conceptions of value and practical reasoning. It is a short step from here to the consistent nonconsequentialism of virtue ethics.

A virtue ethics theory of punishment does provide an adequate account of fault. It portrays fault as an inference, drawn from the particular circumstances and manner of the wrongdoing of the accused, to the effect that the practical judgment of the accused is inadequate or flawed. This defective practical reasoning justifies the incapacitation of the guilty person, in proportion to the defect as it is manifested in minor or major wrongdoing. The adjudication of fault is a prerequisite to just punishment precisely because punishment's justification in any individual case rests on the defect in practical judgment that the inference of fault reveals.

As a final note of introduction, before I make the foregoing arguments in more detail, let me point out one critical contrast between deterrence theory and a virtue ethics theory of punishment. Consequentialism's emphasis on the production of states


27. See, e.g., Oliver Wendell Holmes, Jr., *The Path of the Law*, 10 HARV. L. REV. 457, 459 (1897) ("If you want to know the law and nothing else, you must look at it as a bad man, who cares only for the material consequences which such knowledge enables him to predict. . . .").

28. See infra notes 325-31 and accompanying text.
of affairs in which social welfare is optimized leads the deterrence theorist to overlook the criminal law's concern with the particularities of individual cases.\(^2^9\) Not coincidentally, the particularities of individual cases are the special concern of those features of the criminal law that deterrence theory chronically neglects: fault and adjudication. One can only marvel at the long dominance of a theory of punishment that, at best, finesse the task of explaining two of the criminal law's essential features. My hope is that the reader's appreciation of the magnitude of these omissions will motivate her to see through to the end the present, lengthy exploration of an alternative theory of punishment.

Part I of this Article lays out the basic concepts of the theory of punishment, and describes the retribution and deterrence theories, with an emphasis on the latter. Part II argues that deterrence theory's adoption and adaptation of retributive theory's construction of fault is a failure because deterrence theorists have not developed the notion of nonintentional fault. The best version of deterrence theory moves away from the concept of fault altogether, even though its author, H.L.A. Hart, offered elsewhere good reasons to move in the opposite direction. Part III examines Kahan's "new path" version of deterrence theory and Lessig's "social meaning" version of law and economics. The main burden of this Part is the introduction of the features of nonconsequential value and practical reasoning that Lessig and Kahan fail to address persuasively. Part IV begins with a description of the normative theory that is implied by these features of nonconsequential value and practical reasoning: virtue ethics. The second section of Part IV draws these several threads together in a brief but comprehensive statement of a virtue ethics theory of punishment that explains the justification of punishment, the nature of criminal wrongdoing, and the meaning of fault in the criminal law.

\(^{29}\) See Brudner, supra note 10, at 40-45.
I. THE BASIC CONCEPTS OF THE THEORY OF PUNISHMENT

An adequate theory of punishment should give an account of the two different sets of norms that make up a criminal justice system: the primary norms that govern conduct ex ante and that are addressed to the public at large, and the secondary norms that govern the adjudication of cases ex post and that are addressed to legal authorities. The distinction between primary norms and secondary norms corresponds roughly to the distinction between wrongdoing and culpability, respectively. “Wrongdoing” refers to a violation of the primary norms. A single norm might be stated, in part, as a prohibition and, in part, as a justification; for example, murder is wrongful only if it is not done in self-defense. “Culpability” is a matter of secondary norms, because it is a prerequisite to the imposition of just punishment by legal authorities. The absence of culpability implies either a failure of proof or an excuse. As these alternatives suggest, culpability has two distinct components that are not always separated clearly.


31. See, e.g., FLETCHER, supra note 30, at 491-92 (assigning wrongdoing to primary norms and attribution or culpability to secondary norms); see also HART, supra note 7, at 13-14, 17 (discussing the difference between justification and excuses); HART, supra note 14, at 28, 39 (describing criminal prohibitions as a legislative question and excuses as a judicial question). But see PAUL H. ROBINSON, CRIMINAL LAW § 1.5, at 50 (1997) (arguing that the functions of rule articulation, liability assignment, and grading of offenses are not reflected adequately in these categories).

32. See FLETCHER, supra note 30, at 491-92.

33. See id. at 492.

34. For reasons given in the following text and its accompanying notes, I use the terms “eligibility” and “fault” to describe what I see as the two distinct components of culpability. This terminology is maintained throughout this Article, except where the discussion of another person’s analysis of these issues necessitates the use of that person’s terminology.
First, the accused must have been capable of the alleged wrongdoing, in the sense that he had power or control over the actions in question. The accused is not capable in this sense, and is not eligible for punishment if he acted involuntarily, if he was constrained to act as he did, or if he lacked a minimum level of mental capacity. A presumption exists that the accused is capable and therefore eligible for punishment, and he bears the burden of proving otherwise. Eligibility based on capability is a necessary condition for the imposition of punishment, but not, of course, a sufficient condition. The term "excuse" denotes ineligibility for punishment on grounds of incapability.

A second distinct component of culpability is the requirement that the defendant must be at fault. To find fault is to say that the accused, because of his conduct, deserves blame and punishment. A presumption exists that the accused is not at fault, and the state bears the burden of proving otherwise. In other words, fault, unlike eligibility, is an element of the criminal offense. The absence of fault results in a failure of proof, rather than an excuse.
Whereas eligibility is a necessary condition for punishment and therefore operates unambiguously and uncontroversially to limit criminal liability, the role of fault is more complex. Whether fault is a necessary condition for punishment and whether fault might be a sufficient condition for punishment are questions that define the two orthodox schools of thought in the theory of punishment.

Theories of the justification of punishment commonly are said to fall into two camps: retribution and deterrence. Retributivism is sometimes described inaccurately as the position that fault is a sufficient condition for punishment. More accurately, the retributivist holds that fault is an affirmative, justifying reason to punish: in a case of wrongdoing by a person of ordinary capabilities, fault is a distinct reason to punish that person and also a reason to consider the failure to punish her to be a failure to do justice. This emphasis on fault provides an explanation for what is otherwise a surprising and problematic feature of the criminal law: its relative lack of concern with harm. We punish wrongdoing that does not result in harm, we punish the

the negation of a required fault element of the offense).

42. See supra note 6.

43. See, e.g., MICHAEL MOORE, PLACING BLAME: A GENERAL THEORY OF THE CRIMINAL LAW 193 (1997) ("Culpability, in other words, is sufficient [for punishment], meaning that wrongdoing is not necessary."); R.A. Duff, Penal Communications: Recent Work in the Philosophy of Punishment, in 20 CRIME AND JUSTICE: A REVIEW OF RESEARCH 1, 7 (Michael Tonry ed., 1996) ("My concern here, however, is with positive forms of retributivism that portray criminal desert not merely as a necessary condition of justified punishment (as negative retributivists portray it) but as a sufficient condition... "). It is inaccurate on Duff's part to say that fault (or desert) is a sufficient condition for punishment, because wrongdoing and eligibility for punishment are also necessary conditions. It is inaccurate on Moore's part to say that culpability is sufficient, even assuming that he means culpability to embrace both fault and eligibility, see MOORE, supra, at 404, because wrongdoing is also a necessary condition for punishment. Moore seems to be undecided on this latter point. Compare id. at 193 ("Culpability is necessary to desert, but wrongdoing is not."), with id. at 33 ("To deserve punishment, two things are necessary: one must have done a wrongful action, and one must have done so culpably.").

44. For example, to inject a life-saving drug into a patient who does not wish to receive the drug is a criminal assault, even though the patient's interests are advanced; but to render another football player a quadriplegic with a hard tackle is not a crime, even though he suffers enormous harm. See Brudner, supra note 10, at 21-24 (arguing through this and other examples that the categories of criminal wrongdoing and the infliction of harm do not coincide).

45. See, e.g., MODEL PENAL CODE § 250.10 (1985) ("Except as authorized by law, a
infliction of harm only when we can also blame the actor; and we punish the disposition to do harm when actual harm is only a remote possibility. Generally speaking, this concern with fault, rather than with harm, predominates in the retributivist theory because retributivism conceives of punishment as a matter of duty or principle, and not as a means to promote social welfare.

The principal shortcoming of the retributive theory is that the source and scope of the duty to punish is not clear. Retributivism takes punishment to be required by a kind of fitness in retrospect—a matter of balance or proportion between past and future. It is difficult, however, to articulate the nature of this retrospective fitness, proportion, or balance. Hegel argued that punishment is a matter of bringing the principle of the guilty person's action down on his own head, but this highly abstract argument is a difficult and perhaps flawed chapter in a phi-
sophical tradition that has only recently begun to have an impact on Anglo-American jurisprudence.\textsuperscript{52} Herbert Morris and others have argued that the purpose of punishment is to restore a balance of social benefits and burdens.\textsuperscript{53} But if so, how are we to commensurate the criminal's suffering and the unfair advantage his crime represents?\textsuperscript{54}

The deterrence theory of punishment arguably does not suffer from these burdens because it rests on a less abstract ethical theory: consequentialism. Under consequentialist ethics, punishment is justified by the deterrence of harm or, more broadly, by.

\begin{quote}
RIGHT, supra note 50). The criminal is one who intentionally exercises a degree of freedom that is inconsistent with the freedom of another; for example, by taking the other's goods or by using the other's body as he, the criminal, wills. The criminal act is both an assertion of the criminal actor's status as an absolute end and a denial of that status in the other. This is contradictory because one's status as an absolute end exists objectively only in the free recognition of one's claims of right by other persons who must be recognized as absolute ends in order for their recognition of one's own claims to have any meaning. \textit{See} Brudner, supra note 10, at 32. The retributive model of punishment conceives of punishment as a deprivation of liberty and a denial of end-status, which brings the criminal's principle of action down on his own head. Punishment is "the self-destructive implication of the claimed end-status of the isolated will and therewith the invalidation of that claim; concomitantly, it is the objective confirmation of intersubjectivity as the sole coherent foundation of the person's right to respect." \textit{Id.}

As Brudner points out, however, this conception of criminal wrongdoing is flawed. It purports to premise legal obligation on an unconditional foundation: the formal characteristics of pure will. \textit{See} \textit{id.} at 30. Punishment is supposed to be a necessary implication of the human capacity for choice that defines the person as an absolute end, and a necessary implication of the criminal's particular autonomy-denying exercise of will, as described above. The criminal does not exercise a freedom that is inconsistent with the victim's autonomy, however. The criminal interferes with the embodiments of the victim's autonomous will—his body, his property, etc.—and not with the victim's will itself. The victim's will remains free to renounce his body and property when they are threatened, and the criminal's success over the victim's body or property is not success over the victim's will. Consequently, the foundations of punishment are not unconditional or categorical, but instead are contingent, just as the embodiments of the will are contingent. If the claim of the retributive theory is to premise punishment on a categorical basis, the theory fails. \textit{See} \textit{id.} at 37-38.


54. \textit{See} Duff, supra note 43, at 27 (summarizing this and five other cogent criticisms of the unfair advantage version of retributivism).
the promotion of social welfare.\textsuperscript{55} Deterrence theories of punishment have a strong affinity with the economic analysis of law, because neoclassical economics is a variety of consequentialism. The deterrence theory of punishment is probably dominant over retributivism in criminal law scholarship, though not as much as it used to be,\textsuperscript{56} because consequentialism, and economics in particular, has greater credibility in our society than any comprehensive morality that might undergird a retributive theory.

The deterrence theory of punishment need not take fault to be either a sufficient condition or a necessary condition of punishment. The wholly prospective orientation of consequential justification\textsuperscript{57} means that the justification for punishment need have nothing to do with a past event, such as an actual crime. Similarly, the fact that consequentialism conceives of value in aggregative, agent-neutral terms,\textsuperscript{58} such as wealth or general utility, means that punishment need have nothing to do with an actual criminal. It follows that the deterrence theorist could advocate a system of scapegoating. Provided that the punishment of any randomly selected individual were connected to criminal conduct from the public's point of view, this punishment would promote social welfare by its general deterrent effects.\textsuperscript{59} If deterrence theory does not necessarily limit punishment to actual crimes committed by actual criminals, then, a fortiori, fault can be

\textsuperscript{55} See John Rawls, Two Concepts of Rules, 64 Phil. Rev. 3 (1955), reprinted in THE PHILOSOPHY OF PUNISHMENT: A COLLECTION OF PAPERS, supra note 39, at 105, 107 ("If punishment can be shown to promote effectively the interest of society it is justifiable, otherwise is not.").

\textsuperscript{56} See Duff, supra note 43, at 9-45 (describing the arguments against consequentialism and in favor of retributivism during the latter's revival in the 1970s).

\textsuperscript{57} See Rawls, supra note 55, at 105, 107 ("What we may call the utilitarian view holds that on the principle that bygones are bygones and that only future consequences are material to present decisions, punishment is justifiable only by reference to the probable consequences of maintaining it as one of the devices of the social order.").

\textsuperscript{58} See Philippa Foot, Utilitarianism and the Virtues, 94 Mind 196, 202 (1985) (criticizing the utilitarian concept of an impersonal best state of affairs).

\textsuperscript{59} See J.D. Mabbott, Punishment, 48 Mind 152 (1939), reprinted in THE PHILOSOPHY OF PUNISHMENT: A COLLECTION OF PAPERS, supra note 39, at 39; see also Rawls, supra note 55, at 112 (quoting E.F. Carritt, ETHICAL AND POLITICAL THINKING 65 (1947)) (summarizing this argument against consequentialism).
dispensed with as a necessary condition for just punishment under such a theory.

II. DETERRENCE AND FAULT

Needless to say, no serious deterrence theorist ever has advocated a system of scapegoating. Only a crude form of consequentialism implies the abandonment of wrongdoing as a condition of punishment.\textsuperscript{60} Nor do deterrence theorists fail to find a role for eligibility in the criminal law, because classical political liberalism and the claims of fundamental fairness clearly imply such a limitation.\textsuperscript{61}

Fault is a different story. The concept of fault, and in particular the notion of mens rea—the conception of fault as an intentional state\textsuperscript{62} on the occasion of wrongdoing—is an essentially retributivist idea.\textsuperscript{63} Deterrence theory differs on this point. If the promotion of social welfare is the justifying purpose of punishment, then punishment need not turn on any particular aspect of the wrongdoer or her wrongdoing. Consequentialism authorizes not only nonintentional criminal liability, such as negligence, but also nonfault, or strict, criminal liability.\textsuperscript{64} Furthermore, to

\textsuperscript{60} Rawls argues that a scapegoating system of "telishment"—which did not limit punishment to deprivations imposed by legal authorities under a regular legal process and because of a violation of legal rules—would not be justified by utilitarianism. See Rawls, supra note 55, at 111-12.

\textsuperscript{61} See, e.g., Hart, supra note 14, at 49 (describing fairness and individual liberty as reasons to require capacity and opportunity to conform as prerequisites to just punishment).

\textsuperscript{62} The term "intentional state" refers to the consciousness of an object of attention. See Franz Brentano, Psychology From an Empirical Standpoint 88-89 (Coskar Kraus & Lynda L. McAlister eds., Antos C. Rancurello et al. trans., Humanities Press 1973) (1874) (defining intentionality as "direction toward an object (which is not to be understood here as meaning a thing) . . . ") (footnotes omitted).

\textsuperscript{63} See Brudner, supra note 10, at 33-34 (describing Hegelian retributive theory under which one's intentional wrongdoing denotes a refusal to recognize other persons' status as ends, thereby justifying the same treatment of oneself).

\textsuperscript{64} Hart writes that

\textsuperscript{[s]trict liability is held in some considerable odium by most academic writers and by many judges. But why? What is so precious in the normal requirements of mens rea and how does this normal requirement fit together with our general aims in punishing? It is at this point that scepticism about the old idea that the primary measure of punishment is the wickedness of the criminal act leads to further scepticism about the
the extent that fault has been construed as a moral "wickedness," which is the centerpiece of a crude system of revenge, the elimination of fault from the criminal law has been a goal of the reformist impulse in deterrence theory.

Surprisingly, many deterrence theorists, beginning with Bentham, have sought to retain fault as a feature of punishment, and to retain, moreover, the intentionalist construction of fault. This effort is misconceived, but then so is the effort on the part of other deterrence theorists—notably H.L.A. Hart—to eliminate the concept of fault from the theory of punishment

importance of a responsible act as a condition of punishment . . . . [T]he case is altered if we no longer justify punishment [as retribution or denunciation]; if we think of it as justified by its social aims and effects in protecting society and reforming the criminal . . . .


65. See id. Consider also K.G. Armstrong's ironic portrayal of the view of retributivism in intellectual circles at midcentury:

Retributive punishment is only a polite name for revenge; it is vindictive, inhumane, barbarous, and immoral. Such an infliction of pain-for-pain's-sake harms the person who suffers the pain, the person who inflicts it, and the society which permits it; everybody loses, which brings out its essential pointlessness . . . . The only people who today defend the retributive theory are those who, whether they know it or not, get pleasure and a feeling of virtue from seeing others suffer, or those who have a hidden theological axe to grind.


66. Rawls points out that "[h]istorically, [utilitarianism] is a protest against the indiscriminate and ineffective use of the criminal law. It seeks to dissuade us from assigning to penal institutions the improper, if not sacrilegious, task of matching suffering with moral turpitude." Rawls, supra note 55, at 110 (citing LEON RADZINOWICZ, A HISTORY OF ENGLISH CRIMINAL LAW AND ITS ADMINISTRATION FROM 1750: THE MOVEMENT FOR REFORM 1750-1833 (1948)).

67. See, e.g., HART, supra note 64, at 181 ("Such a reinterpretation [of responsibility] would not stress, as our legal moralists do, the importance of judgments of degrees of wickedness about which there is far less agreement than they suppose."). Hart's argument here, as in many of the essays published in Punishment and Responsibility, is that eligibility considerations are a sufficient limitation on punishment that is justified by social welfare. See infra notes 120-52 and accompanying text.

68. See JEREMY BENTHAM, AN INTRODUCTION TO THE PRINCIPLES OF MORALS AND LEGISLATION 82-88, 170-77 (Hafner Pub'l'g Co. 1970) (1780).

69. See infra notes 106-19 and accompanying text.
altogether.\footnote{See infra notes 120-52 and accompanying text.} Just punishment is not confined to cases of intentional wrongdoing, but not for reasons that a consequentialist would give.

The central failure of both strains of deterrence theory involves the inability to develop a conception of nonintentional fault. This task is one that Kahan's "new path" deterrence theory undertakes;\footnote{See infra notes 181-291 and accompanying text.} but Kahan embarks on this project almost unwittingly and, not surprisingly, without success. Only if we place the conceptual odds and ends that Kahan has assembled into a coherent ethical theory and then bring this theory to bear on the problems of the criminal law can we see a genuinely new path in the theory of punishment.

A. Deterrence and the Intentionalist Construction of Fault

1. Fault as a Distribution Rule

The deterrence theorist's answer to the scapegoating objection turns on a distinction between a practice taken as a whole, and an instance or application of the practice—a distinction which was critically important to deterrence theory at midcentury,\footnote{See HART, supra note 7, at 5-11 (responding to "the stock 'retributive' argument" by means of the same distinction); Rawls, supra note 55, at 111; see also Mabbott, supra note 59, at 50 ("One may consider the merits of a legal system . . . but the acceptance of [these merits] involve[] the surrender of utilitarian considerations in particular cases as they arise."); Anthony M. Quinton, On Punishment, 14 ANALYSIS (1954), reprinted in THE PHILOSOPHY OF PUNISHMENT: A COLLECTION OF PAPERS, supra note 39, at 55, 63.} and which Rawls traced to Hume.\footnote{See Rawls, supra note 55, at 105 & n.2 (citing DAVID HUME, A TREATISE OF HUMAN NATURE bk. III, pt. ii, §§ 2-4).} Deterrence theorists argue that the practice of punishment—by which they mean the criminal code—can be justified in consequentialist terms. Instances of the practice—that is, punishment in individual cases—can be justified on a different basis; for example, by legal rules that employ traditional categories of desert and fault.\footnote{Rawls puts the point this way:

[Al] particular man in punished, rather than some other man, because he is guilty, and he is guilty because he broke the law (past tense). In his case the law looks back, the judge looks back, the jury
which includes within it a set of desert-based rules about who shall and shall not be punished will not tolerate or dictate scapegoating, even if the criminal code as a whole is justified on consequentialist grounds.

Hart makes an argument like the one in the preceding paragraph, and does so in terms that might lead a careless reader to think that he is advancing a genuinely mixed theory of punishment, in which the justification of the institution of punishment is accomplished by consequentialism, and the justification of the infliction of punishment in individual cases is accomplished by a deontological morality. But this is not quite right. Hart quite studiously avoids the term “justification of punishment” where individual cases are concerned, in favor of the term “distribution of punishment.” He employs the practice/instance of a practice distinction to confine the question of extralegal, moral justification to the institution of punishment as a whole. Individual cases of punishment are justified only legally, not morally, within the institution of punishment, by means of legal rules for the distribution of punishment. Legal rules for the distribution of

looks back, and a penalty is visited upon him for something he did. That a man is to be punished, and what his punishment is to be, is settled by its being shown that he broke the law and that the law assigns that penalty for the violation of it.

On the other hand we have the institution of punishment itself, and recommend and accept various changes in it, because it is thought by the (ideal) legislator and by those to whom the law applies that, as part of a system of law impartially applied from case to case arising under it, it will have the consequence, in the long run, of furthering the interests of society.

One can say, then, that the judge and the legislator stand in different positions and look in different directions: one to the past, the other to the future. The justification of what the judge does, qua judge, sounds like the retributive view; the justification of what the (ideal) legislator does, qua legislator, sounds like the utilitarian view.

Rawls, supra note 55, at 108; see also HART, supra note 14, at 39 (making the same distinction).

75. Anthony Quinton's essay, On Punishment, makes it clear that the principal effect of the distinction between the justification of a practice and of instances of a practice is to deprive fault or desert, or, in Quinton's terminology, "guilt," of any connection to an extralegal morality, including utilitarianism, which requires punishment.

The distinction between setting up and applying penal rules helps to explain the different parts played by utility and guilt in the justification
punishment might be written in terms of desert, but they do not incorporate deontological morality.

The deterrence theorist's desert-based legal rules for the distribution of punishment reflect a rump retributivism, at best. John Rawls contends that utilitarians "could agree with Bradley that: 'Punishment is punishment only when it is deserved.'" But Rawls understands deserved punishment to be a deprivation imposed by authorities in a regular legal process, because of a violation of legal rules. The fact that this is not the retributivist's idea of desert is evident from Rawls's citation to Hobbes for this definition of deserved punishment. Hobbes's consequentialism in the matter of punishment is perfectly clear. Desert is implied by fault, and the concept of fault is correspondingly attenuated in this consequentialist system. Hart rejects the use of the word "fault" in this context precisely because it suggests an extralegal, moral justification in individual cases of punishment.

To interpret fault as a legal rule for the distribution of punishment has a second implication, in addition to depriving fault of any role in the moral justification of punishment. Under the deterrence theorist's interpretation of fault as a distribution rule, the absence of fault is a reason not to punish, but the presence of fault ceases to be a reason to consider the failure to punish to be unjust. In other words, fault loses its force as an affirmative of punishment, in particular the fact that where utility is a moral, guilt is a logical, justification. Guilt is irrelevant to the setting up of rules, for until they have been set up the notion of guilt is undefined and without application. Utility is irrelevant to the application of rules, for once the rules have been set up punishment is determined by guilt, once they are seen to apply the rule makes a sentence of punishment necessarily follow.

Quinton, supra note 72, at 63.

76. Rawls, supra note 55, at 109 n.8.

77. Id. at 109-110 (citing THOMAS HOBBES, LEVIATHAN, at ch. xxviii (Richard Tuck ed., Cambridge Univ. Press 1991) (1651)); see also Benn, supra note 49, at 332-34; Quinton, supra note 72, at 63 ("Punishment is the infliction of suffering attached ... to certain kinds of action . . . .").

78. See HOBBES, supra note 77, at 215 ("Fifthly, that all evil which is inflicted without intention, or possibility of disposing the Delinquent, or (by his example) other men, to obey the Lawes, is not Punishment; but an act of hostility; because without such an end, no hurt done is contained under that name.").

79. See HART, supra note 14, at 40.
reason to punish. Fault becomes only a side constraint on punishment. The effect of this interpretation is to align fault functionally with the other culpability concept, eligibility for punishment according to capability. Eligibility concerns, such as those which underwrite the duress and insanity defenses, operate in a wholly liability-limiting way. The distribution rule, side-constraint view of fault places fault in the same role, and provides the deterrence theorist with a unitary conception of culpability, in which the distinction between fault and eligibility is obscured.

One additional implication of the deterrence theorists' response to the scapegoating charge is worth notice here. Having assigned desert and fault to the adjudicative stage of the criminal law, deterrence theorists deem the legislative task to be more fundamental than the task of adjudicating individual cases. The adjudicative task is portrayed as an administrative matter that is devoid of any independent significance. This is a mistake, as we will see. It is a mistake that was perhaps more apparent when the majority of cases were not disposed of by plea bargaining. But of course the current dominance of plea bargaining might well be attributed, at least in part, to the dominance of deterrence theory and its implicit depreciation of the criminal trial.

2. Fault, Distribution Rules, and Intentions

Bentham failed to perceive any conflict between consequentialism and the requirement of mens rea—the essentially retributivist construction of fault as an intentional state on the occasion of wrongdoing—because his rationalistic conception of

---

80. For example, Rawls argues:

One might say, however, that the utilitarian view is more fundamental since it applies to a more fundamental office, for the judge carries out the legislator's will so far as he can determine it. Once the legislator decides to have laws and to assign penalties for their violation (as things are there must be both the law and the penalty) an institution is set up which involves a retributive conception of particular cases.

Rawls, supra note 55, at 108.

81. See HART, supra note 7, at 7 ("Yet only if we keep alive the distinction... between the primary objective of the law in encouraging or discouraging certain kinds of behaviour, and its merely ancillary sanction or remedial steps, can we give sense to the notion of a crime or offence.").
deterrence led him to believe that nonintentional wrongdoing, including negligence, could not be deterred in any case.\(^\text{82}\) He argued that a criminal prohibition could not and should not operate on the potential criminal "[i]n the case of un intentionality; where he intends not to engage, and thereby knows not that he is about to engage, in the act in which eventually he is about to engage."\(^\text{83}\) Hart argued in response to Bentham that the claim that inadvertent wrongdoing cannot be deterred rests on "a spectacular non sequitur."\(^\text{84}\) Both specific and general deterrence clearly can be served by the imposition of liability for negligence. Because negligence is not only inadvertence, but also—and more to the point—the failure to meet a standard of due care, the imposition of negligence liability deters wrongdoing because it encourages both this defendant and other actors to exercise greater care in the future.\(^\text{85}\) As Hart recognized, not only negligence, but also strict liability (if not absolute liability\(^\text{86}\)) is perfectly consistent with consequentialism. If aggregate social welfare is the standard of right action, the distribution rules which will optimize that value may or may not have anything to do with intentional states.

Nevertheless, the predominant tendency in deterrence theory has been to write rules for the distribution of punishment in terms of intentional states. The principal instance of this pattern is section 2.02 of the Model Penal Code. Section 2.02's system of three intentional states,\(^\text{87}\) which are to be proved in con-

---

82. See Bentham, supra note 68, at 172-75.
83. Id. at 174. Glanville Williams advanced this rationale for excluding negligence as a kind of fault during the drafting of the Model Penal Code. See MODEL PENAL CODE § 2.02 cmt. 4 (1985) (citing GLANVILLE WILLIAMS, CRIMINAL LAW: THE GENERAL PART (2d ed. 1961)); see also Williams, supra, at 123 (arguing that the threat of punishment for negligence must be ineffective because the negligent actor "does not realise that it is addressed to him").
84. HART, supra note 7, at 19.
85. See MODEL PENAL CODE § 2.02 cmt. 4; H.L.A. HART, INTENTION AND PUNISHMENT, IN PUNISHMENT AND RESPONSIBILITY: ESSAYS IN THE PHILOSOPHY OF LAW, supra note 7, at 113, 132-35.
86. Absolute liability would not recognize excusing conditions, in addition to permitting criminal liability without proof of fault. See Hart, supra note 13, at 149-57.
87. See MODEL PENAL CODE § 2.02(2) (defining purpose, knowledge, and recklessness as "Kinds of Culpability"). The mental states of purpose, knowledge, and recklessness are all defined by reference to harm or a risk of harm of which the offender
nection with three categories of offense elements\textsuperscript{88} constitutes a set of legal rules on fault. Section 2.02 lists negligence, a nonintentional state, as a "kind of culpability," but subsection 2.02(3) states a presumption in favor of intentional fault.\textsuperscript{89} and the Code's commentary makes it clear that negligence liability is in all cases disfavored as a normative matter.\textsuperscript{90} Whereas the retributivist limits criminal liability to cases of intentional wrongdoing because inadvertent wrongdoing, including negligence, does not trigger the rationale of punishment,\textsuperscript{91} the consequentialist deterrence theorist has different reasons to draw the line, at least presumptively, at intentional wrongdoing. One can see at least three such reasons.

The first reason is simply a basic requirement of any plausible legal theory: it must generate a recognizable description of existing law. This is the motive behind the recognition of fault even in the truncated form of desert-based distribution rules, and the adoption of intentional states as the organizing principle of the legal rules for the distribution of punishment seems to follow as a matter of course. Hart often seems inclined to articulate a sys-

\textsuperscript{88} See MODEL PENAL CODE § 2.02 cmt. 1 (distinguishing circumstance, conduct, and result elements for this purpose).

\textsuperscript{89} The Model Penal Code's "elements approach" to fault requires that a level of fault shall be proved in conjunction with each material element of every Code offense. See MODEL PENAL CODE § 2.02(1); see also Paul H. Robinson & Jane A. Grall, Element Analysis in Defining Criminal Liability: The Model Penal Code and Beyond, 35 STAN. L. REV. 681, 683 (1983) (stating that jurisdictions that have adopted the Model Penal Code must "apply an element analysis to each offense and theory of liability"). Under section 2.02(3) of the Code, the presumptive level of fault for each and every element in the Code is recklessness. Recklessness is defined in intentional-state terms as a conscious disregard of a substantial risk of harm. See MODEL PENAL CODE § 2.02(2)(c).

\textsuperscript{90} See, e.g., MODEL PENAL CODE § 2.02 cmt. 4 ("Accordingly, negligence, as here defined, should not be wholly rejected as a ground of culpability that may suffice for purposes of penal law, though it should properly not generally be deemed sufficient in the definition of specific crimes and it should often be differentiated from conduct involving higher culpability for purposes of sentence.") (citations omitted).

\textsuperscript{91} Whereas intentional wrongdoing implies a refusal to recognize other persons' status as ends, thereby justifying the same treatment of oneself, inadvertent wrongdoing does not have this implication. See Brudner, supra note 10, at 33-34.
tem of universal strict liability, 92 but there would be little point in his doing so. The criminal law has never operated in such a way, and it is not likely to do so in the foreseeable future. The majority of offenses do contain mens rea elements, and these must be given some kind of recognition even by as committed a consequentialist as Hart.

Second, consequentialism and consequentialists from Bentham and Mill onward have been associated with classical political liberalism. 93 This association has carried over into the deterrence theory of punishment. 94 For example, the deterrent theorist's rejection of fault as an affirmative reason to punish, and his requirement of harm instead, 95 is an application of the harm principle which is a central feature of political liberalism: the notion that government should not coerce individuals except in order to prevent them from harming one another. 96 The requirement of fault as a necessary condition on just punishment can also serve the liberal ideal of limited government, and deterrence theorists typically attribute this function to it. 97 Here, too, the Model Penal Code's section 2.02 is the principal example. Consider the comments of one enthusiastic commentator:

The drafters of the Model Penal Code achieved a monumental advance in the law through the formulation of precisely defined mental states and the application of those mental states to other components of penal offenses. They have accorded us a vocabulary and a grammar with which we

92. See HART, supra note 64, at 176-77 (arguing that strict liability might best promote social welfare).
94. See, e.g., HART, supra note 14, at 48-49 (culpability concepts serve to maximize choice and respect individual dignity).
97. See, e.g., HERBERT L. PACKER, THE LIMITS OF THE CRIMINAL SANCTION 112 (1968) ("Our best reason for not punishing people who are not 'culpable' is simply that culpability is an appropriate criterion for limiting the reach of state intervention.").
may reach and express a conclusion through an ordered thought process that can be replicated by others. We no longer need to depend upon the mixture of moral precepts and emotional interplays that underlaid our previous assessment of mens rea—an evil that we tolerated for centuries, in part because of our inability to recognize it, and in part because of our inability to exorcise it.98

The tendency of the distinct philosophical doctrines of liberalism, consequentialism, and legal positivism to reinforce one another is on full display in a passage such as this one.

Third, the requirement of an intentional state in the proof of a criminal offense contributes to the attainment of optimal deterrence from the economist's point of view.99 Criminal penalties are sanctions that do not correspond to the actual costs of violations of the law's conduct rules; the typical penalty imposes a greater cost than the cost of the violation.100 This discrepancy results in overdeterrence: people engage in less than an optimal amount of the activity punished.101 Overdeterrence of crime in the sense that we will have a sub-optimal amount of it is not really a problem, because the utility of crimes is illicit: we want no murders, not an optimal number of murders.102 However, if

100. See Cooter, supra note 99, at 1532-33, 1548-50. Hobbes argued that a deprivation that did not have this characteristic was not punishment. See HOBSES, supra note 77, at 215 ("Seventhly, If the harm inflicted be lesse than the benefit, or contentment that naturally followeth the crime committed, that harm is not within the definition; and is rather the Price, or Redemption, than the Punishment of a Crime. . . .").
101. See Cooter, supra note 99, at 1550; see also RICHARD A. POSNER, ECONOMIC ANALYSIS OF LAW 243-44 (5th ed. 1998) (describing expected punishment cost versus costs of enforcement); Shavell, supra note 99, at 1234-36 (describing deterrence as a situation in which the "expected sanction" exceeds the "expected private benefits"); George J. Stigler, The Optimum Enforcement of Laws, 78 J. POL. ECON. 526, 527 (1970) (stating that "increasing the punishment would seem always to increase the deterrence").
102. See Stigler, supra note 101, at 527.
we are concerned that we might also overdeter legitimate activity, then it is clear that we should not premise criminal liability on negligence. Negligence is not only a secondary norm; it is also a primary norm, a requirement that we recognize some risks. But the content of the negligence norm—the particular risk that the accused ought to have recognized—is specified retrospectively, at the adjudicative stage of the criminal justice process, rather than prospectively, at the legislative stage. This ex post primary-normativity leaves people who are situated ex ante in a state of uncertainty, and overdeterrence is almost certain to follow. This threat of overdeterrence is a good consequentialist reason to limit criminal liability to cases of intentional wrongdoing.

3. The Problem of Nonintentional Fault

The difficulty with the consequentialist interpretation of fault as a side constraint on punishment under an intentionalist construction is that fault is not in fact confined to intentional wrongdoing—but not for reasons which a consequentialist would give.

Instances abound of genuine nonintentional fault that authorizes severe punishments for serious crimes—cases of unreasonable mistake, willful ignorance, voluntary intoxication, felony murder, "transferred" intentions, and depraved indifference. The orthodox view—fault consists only of proof of intentional states as a side constraint on punishment—relies heavily on the denial or minimization of these contrary instances of nonintentional fault. This is so not least because these instances of fault undermine the midcentury deterrence theorists’ reformist construction of fault as a violation of legal rules that has no connection to extralegal morality. Nonintentional fault is genuine fault—genuine in the sense that it is an affirmative, justifying reason to pun-

103. See HART, supra note 13, at 152-56 (describing the distinct conduct-governing and liability-limiting components of negligence).

104. See, e.g., POSNER, supra note 101, at 243-44 (describing the costs of overdeterrence that result from criminal negligence liability and the role of intentionality in reducing them); see also Parker, supra note 99, at 759-60 (confining criminal liability to intentional wrongdoing avoids excessive costs of "self-characterizing behavior," i.e., determining whether or not one is in compliance with an indeterminate standard).
ish—but consequentialist theorists deny the existence of any such thing.

Consider a prominent recent example of the orthodox approach that is among those theories Kahan has spotted on the new path of deterrence. Paul Robinson and John Darley offer a deterrence-based rationale in favor of limiting fault to intentional states and in favor of rejecting or limiting negligence. Robinson and Darley argue that the criminal law cannot effectively deter wrongdoing unless it reflects popular intuitions about desert. They endorse the Model Penal Code's presumption in favor of intentional fault under section 2.02(3) on the ground that such a presumption is reflected in their empirical studies of popular intuitions about desert. Specifically, they claim to have discovered an intuition that crimes of negligence are less serious and should be punished less severely than crimes of recklessness—an intuition that corresponds to the intentionalist scheme of subsection 2.02(3).

The proposition that subsection 2.02(3)'s intentionalist construction of fault invariably reflects our intuitions about desert is highly questionable. Contrary intuitions are easy to come by in the criminal law. Consider one instance of a contrary intuition regarding nonintentional fault.

In cases of mistake of fact, the intentionalist construction of fault is demonstrably inadequate because it implies wildly implausible acquittals. If the accused has made a mistake regarding a material element of a crime, then he has no intentional state regarding that element. Because this absence results in a failure of proof on that element, the mistake of fact should result in an acquittal. But consider this: If the nonconsent of a rape

105. See Kahan, supra note 1, at 2481 (citing Paul H. Robinson & John M. Darley, The Utility of Desert, 91 NW. L. REV. 453 (1997)).
107. See Robinson & Darley, supra note 105, at 477-78.
109. See id. at 87.
110. See id.; see also supra note 89 and accompanying text (describing section 2.02(3)'s intentionalist construction of fault).
111. See MODEL PENAL CODE § 2.02 cmt. 1 (1985) (describing the necessity of proof of some level of culpability with regard to each and every material element of an of-
victim is an element of rape, then a mistake regarding nonconsent will acquit the rapist. On the intentionalist construction of fault the accused rapist is not at fault because he had no intentional state regarding nonconsent. Furthermore, an unreasonable mistake, no matter how outrageous, will have the same implication: acquittal for a failure of proof. 112 To require the mistake regarding consent to be reasonable in order to acquit the defendant would be to impose liability for rape on proof of no more than negligence; it would premise punishment on the proposition that the accused should have recognized the victim's nonconsent and refrained from intercourse because a reasonable man would have done so. 113 On the intentionalist construction of fault, however, negligence can premise only low-level punishments for minor crimes 114—certainly nothing approaching the magnitude of rape.

Robinson appears to insist on the propriety of this counterintuitive result 115—acquittal in cases of rape for outrageously unreasonable mistakes regarding the victim's nonconsent—in the face of nearly total defection on this point by jurisdictions that have adopted some version of the Code. Unlike the Code's drafters, legislatures and courts have appreciated the fact that an offense such as rape might involve a kind of fault that cannot be reduced to a discrete mental state. 116 Their solution to the

---


114. See MODEL PENAL CODE § 2.02 cmt. 4 (cautioning against the use of negligence in the definition of offenses and contrasting it with "higher" culpability levels).

115. See 1 ROBINSON, supra note 36, § 62(c)(2), at 253.

116. For example, suppose that a defendant claims that he genuinely believed that the victim of his rape, his friend's wife, consented to have sex with him. The defen-
problem is instructive. They have adopted a reasonableness requirement for mistake of fact in cases of rape.\textsuperscript{117} The intentionalist objects that a reasonableness requirement in this context improperly premises rape on negligence.\textsuperscript{118} However, if this is so, then Robinson and Darley are mistaken in their claim that popular intuitions about desert place negligence at the low end of the fault spectrum and limit its role to minor crimes and punishments.

As it happens, Robinson and Darley are not so much mistaken in their account of negligence as they are mistaken in their defendant claims that he and his friend were extremely drunk, and his friend persuaded him that the friend's wife was ""kinky"" and that she would be ""turned on"" by being raped. In light of these representations, the defendant interpreted the victim's overt refusals and resistance to intercourse as consent. On an intentionalist construction of mistake and fault, this defendant is entitled to an instruction that the proof of rape will fail in this case, because (if the jury believes his story) the defendant lacked a mental state of purpose, knowledge, or recklessness regarding the victim's nonconsent to intercourse—an essential element of rape. This defendant is nevertheless at fault and his fault lies in the combination of his particular circumstances, which includes his severe voluntary intoxication, his poor choice of friends, his evident ability to make himself believe whatever he finds convenient to believe, and a general moral obtuseness, as evidenced by his failure to perceive not only a woman's genuine resistance to forced sexual intercourse, but also the fact that even a simulated rape is an act degrading to human dignity. If such a defendant is denied a jury instruction setting forth the intentionalist construction of mistake and fault, to affirm his conviction would nevertheless be consistent with reason and justice. See Morgan, [1976] App. Cas. at 203-04 (Lord Cross), 214-15 (Lord Hailsham), 237-39 (Lord Fraser) (affirming rape convictions of several British military officers on these facts) (citing Criminal Appeal Act, 1968, Ch. 19 § 2(1) (Eng.) (authorizing the affirmance of convictions in spite of error when not inconsistent with justice)).

\textsuperscript{117} See KADISH & SCHULHOFER, supra note 6, at 327 ("Most of the recent American cases permit a mistake defense [to rape], but only when the defendant's error as to consent is honest and reasonable."); Harriet R. Galvin, \textit{Shielding Rape Victims in the State and Federal Courts: A Proposal for the Second Decade}, 70 MINN. L. REV. 763, 849 (1986) ("[A] majority of courts here have adopted the rule that only a \textit{reasonable} mistake is a valid defense.") (citing People v. Mayberry, 542 P.2d 1337, 1344-47 (Cal. 1975); State v. Dizon, 390 P.2d 759, 769 (Haw. 1964)); Victoria J. Dettmar, Comment, \textit{Culpable Mistakes in Rape: Eliminating the Defense of Unreasonable Mistake of Fact as to Victim Consent}, 89 DICK. L. REV. 473, 483-89 (1985) (summarizing cases that impose reasonableness requirement on mistake in cases of rape); see also Lani Anne Remick, Comment, \textit{Read Her Lips: An Argument for a Verbal Consent Standard in Rape}, 141 U. PA. L. REV. 1103, 1108 ("At least one court has even gone so far as to suggest that there is no mens rea for rape.").

\textsuperscript{118} See, e.g., 1 ROBINSON, supra note 36, § 62(c)(2), at 253 & n.23 (listing and criticizing reasonable mistake provisions).
understanding of fault generally. The case of the unreasonably mistaken rapist points toward the existence of a kind of fault that is nonintentional, but that also is sufficiently grave to justify severe punishments for serious offenses. This kind of fault does resemble negligence. But, if to premise rape on negligence would be improper, this should tell us, not that an unreasonably mistaken rapist is only minimally at fault—an absurd proposition—but instead that his fault only resembles negligence. It is a different kind or degree of nonintentional fault. Were deterrence theorists able to articulate a theoretical account of the full range of nonintentional fault, it might ease the tension between the expansive tendencies of their consequentialist rationale for punishment and their liberal and efficiency-minded efforts to construe fault as a limitation on criminal liability. Unfortunately, the section of the new path of deterrence that Robinson and Darley have blazed will not take consequentialists this far, or even in the right direction.

In any event, we have been down this path before, and it is a dead end. The Robinson and Darley thesis suffers from a deeper defect. Their thesis implies that the criminal law operates under a pretense that its claims about fault are meaningful, with a wink and a nod to consequentialist theorists and a legal elite who know better. We do not impose criminal liability up to the logical limits of the consequentialist rationale because to do so would risk a popular backlash, but the criminal law, properly speaking, is agnostic about the moral intuitions that would drive this popular backlash. The difficulty with this argument is that its "as if" construction represents one thought too many; it would be more parsimonious to assume instead that we mean what we say in the law and that we always have done so—that the criminal law's claims about fault and desert really are meaningful. H.L.A. Hart suggested this argument against a different version of the same thesis over thirty years ago:

[We do not dissociate ourselves from the principle that it is wrong to punish the hopelessly insane or those who act unintentionally, etc., by treating it as something merely embodied in popular mores to which concessions must be made sometimes. We condemn legal systems where they disregard this principle; whereas we try to educate people out of their pref-
erence for savage penalties even if we might in extreme cases of threatened disorder concede them.\textsuperscript{119}

In this passage, Hart addresses the recognition of excusing conditions—the question of eligibility rather than fault—but his point holds for both culpability concepts. Ironically, Hart himself did not see this.

B. Hart's Rejection of Fault

Hart's account of culpability stands out from those of his fellow consequentialists because he does not argue for the presumptive limitation of criminal liability to cases of intentional wrongdoing. Hart embraces negligence liability and strict liability as natural implications of consequentialism and as acceptable bases of punishment.\textsuperscript{120} He insists only that we do not abandon the necessary condition of eligibility for punishment.\textsuperscript{121} For example, Hart argues that negligence is not merely inadvertence, but is also, and more to the point, a standard of conduct—a primary norm to the effect that one ought to recognize and respond to the risks that attend everyday life.\textsuperscript{122} The imposition of punishment for negligence is not unjust if, as is usually the case, this primary norm is conjoined with a secondary norm to the effect that punishment shall not be inflicted if the accused lacked the capacity to comply with the standard of due care.\textsuperscript{123}

The foregoing analysis of negligence is unassailable on its own terms. Nonetheless, it is important to see that Hart's analysis eliminates any notion of fault from the law. A good case of negligence consists of only the violation of a prohibition by a person who is eligible for punishment. No role exists for the notions of desert and fault; no appeal is made to an affirmative, justifying reason to punish above and beyond the reason provided by the act of wrongdoing. In this respect, Hart's analysis of negligence is paradigmatic of his analysis of culpability generally. Hart recognizes the distinction between fault and eligibility,\textsuperscript{124} but he

\begin{itemize}
\item \textsuperscript{119} HART, supra note 7, at 21.
\item \textsuperscript{120} See HART, supra note 85, at 132.
\item \textsuperscript{121} See id. at 132-35; HART, supra note 64, at 176-85.
\item \textsuperscript{122} See HART, supra note 13, at 147-48.
\item \textsuperscript{123} See id. at 149-52.
\item \textsuperscript{124} Hart writes:
\end{itemize}
explicitly refuses to observe it, and does so "with a clear conscience, since little is to be gained in clarity by a rigid division which the contemporary use of the expression mens rea often ignores." Hart does recognize a liability-limiting function that is associated with the mens rea elements of criminal offenses; but whereas other deterrence theories give these elements a liberal or economic reading, or an "as if" construction such as Robinson's, Hart's analysis incorporates them into the matter of eligibility.

As an illustration of Hart's treatment of fault and eligibility, consider a Hartian analysis of the case of the unreasonably mistaken rapist. As we have seen, the intentionalist construction of fault implies that a mistake of fact can be analyzed as a failure of proof on an element of intentional fault. Hart does not subscribe to this analysis of mistake because he does not endorse the intentionalist construction of fault. In accordance with an older tradition in English law, Hart categorizes mistake among the excusing conditions. Under this theory, only a reasonable mistake will excuse. This means that Hart would have no difficulty convicting the unreasonably mistaken rapist and, in this respect, he seems to capture our intuitions.

On the other hand, Hart does not provide us with an account of nonintentional fault in the case of the unreasonably mistaken rapist, or in any other case. Hart argues forcefully that "moral culpability" is not required for a criminal conviction and that the use of the word "fault" obscures this fact. Hart would explain

Continental codes usually make a firm distinction between these two main types of psychological conditions: questions concerning general capacity are described as matters of responsibility or 'imputability,' whereas questions concerning the presence or absence of knowledge or intention on particular occasions are not described as matters of 'imputability,' but are referred to the topic of 'fault' (Schuld, faute, dolo, & C.).

HART, supra note 15, at 218.
125. Id. at 219.
126. See supra text accompanying note 111.
127. See HART, supra note 13, at 152; see also R. v. Tolson, [1886-90] All E.R. 26, 37 (analyzing mistake as an excuse); Richard H.S. Tur, Subjectivism and Objectivism: Towards Synthesis, in ACTION AND VALUE IN CRIMINAL LAW, supra note 10, at 213, 218 (distinguishing mistake as an excuse under Tolson from the intentionalist construction of mistake as the negation of fault).
128. See HART, supra note 14, at 35-40.
the conviction of the unreasonably mistaken rapist as an instance of the violation of a positive legal prohibition under circumstances not meriting an excuse, in which considerations of public welfare provide ample support for the existence and enforcement of the prohibition. Fault in the retributivist’s sense of a distinct, affirmative, justifying reason to punish for the sake of proportion between past and future, or in any sense that would connect fault to the justifying purposes of punishment, plays no role in Hart’s analysis.

We have ample reasons to reject this treatment of fault. The first reason is an empirical one. Hart insisted on the conflation of fault and eligibility even though he knew that the evidence of the law did not support it. Hart admitted that judges do not analyze questions of fault in the language of eligibility (which Hart referred to as “responsibility”): “I have not succeeded in finding cases where a normal person, merely lacking some ordinary element of knowledge or intention on a particular occasion, is said for that reason not to be responsible for that particular action, even though he is for that reason not liable to punishment.” This passage, however, appears on the same page as Hart’s assertion that “the contemporary use of the expression mens rea” ignores the distinction between fault and eligibility.

Nothing in the intervening text accounts for Hart’s disregard of the evidence that he has just acknowledged. Furthermore, Hart might have noticed that the prosecution carries the burden of proof on mens rea, but not on a defense such as insanity—a

---

I think that the use of the word ‘fault’ in juristic discussion to designate the requirement that liability be excluded by excusing conditions may have blurred the important distinction between the assertions that (1) it is morally permissible to punish only voluntary actions and (2) it is morally permissible to punish only voluntary commission of a moral wrong.

Id. 129. See id.

130. See, e.g., HART, supra note 13, at 152 (describing the capabilities that determine eligibility as “aspects of responsibility”); HART, supra note 15, at 219 (referring to eligibility considerations as the “criteria of responsibility”).


132. Id.

133. Compare Woolmington v. Director of Pub. Prosecutions, [1935] App. Cas. 462, 481 (reciting the duty of the prosecution to prove all elements of the offense charged), with Sodeman v. R., [1936] 2 All E.R. 1138, 1138 (stating that the defen-
practical difference between allegations of fault and claims of eligibility that indicates significant differences between fault and eligibility in theory.

It seems fair to say that Hart's desire, as a liberal reformist, to eliminate the concept of fault from the theory of punishment clouded his view of the evidence in the case law. Hart thought that the terminology of fault imported too robust a conception of desert into the theory of punishment.\textsuperscript{134} He seems not to have considered the possibility that he was quite ready to acknowledge in the matter of eligibility and the excuses: the possibility that the language of fault has genuine meaning; that fault is not "something merely embodied in popular \textit{mores} to which concessions must be made sometimes"\textsuperscript{135} and that simply to read fault out of the law is unjustifiable.

In any event, Hart's conflation of fault and eligibility is not only unsupported by the evidence of the law, it is also unsustainable in theory. Consider Hart's analysis of mens rea. Criminal offense definitions include elements that are intentional states, even in a penal code that is not premised on the view that an intentional state should ordinarily be proved in connection with each and every material element. Hart recognizes these intentional elements—he could hardly do otherwise—under the heading of mens rea.\textsuperscript{136} If Hart's position is that these mental elements do not represent genuine fault, an affirmative, justifying reason to punish, then they must represent only limitations on punishment; they must function only to enhance the burden of proof that the government carries. This view aligns mens rea, functionally with eligibility, and merges mens rea and eligibility into a single, unitary culpability requirement. In Hart's terms, mens rea is not part of the justifying purpose of punishment; it is, like excusing conditions, a matter of the distribution of punishment.\textsuperscript{137}

\begin{itemize}
\item \textsuperscript{134} See \textsc{Hart}, \textit{supra} note 14, at 40.
\item \textsuperscript{135} \textsc{Hart}, \textit{supra} note 7, at 21.
\item \textsuperscript{136} See \textsc{Hart}, \textit{supra} note 85, at 114.
\item \textsuperscript{137} See \textsc{Hart}, \textit{supra} note 7, at 9 (distinguishing the justifying purpose and the distribution of punishment).
\end{itemize}
The difficulty is that this view of mens rea implies a radical and unsupportable separation between the mental elements of a crime and its other material elements. The particular wrongdoing in an offense strongly determines the nature of the mental states that must be proved in connection with the offense. This connection to the primary norm allies these mental states with punishment’s justifying purpose and undermines Hart’s position that they pertain only to punishment’s distribution. Indeed, it undermines the assumption among deterrence theorists generally that the distinction between the justification of a practice and the justification of instances of that practice applies to punishment in a way that will support deterrence theory.  

In *A Plea for Excuses*, Austin famously remarked that excuses are essentially failures of or interferences with normal capacities for action, and so throw light on the conditions of responsibility. However, a second remark in the same passage is of more significance for Austin’s argument as a whole. He writes: “Further, it emerges that not every slip-up occurs in connexion with *every* thing that could be called an ‘action,’ that not every excuse is apt with every verb—far indeed from it. . . .” Excuses are characteristically expressed as adverbs, but they occur in ordinary speech in verb phrases, and not as adverbs standing alone. Adverbs have a limited range of application; that is, not only must they go with a verb, they make sense only with some specific verb or kind of verb. The upshot of these observations is that we cannot understand the reason for an excuse in any given case unless we understand the particular prohibition that is at issue.

---

138. *See supra* text accompanying notes 72-81.  
139. *See* Austin, *supra* note 17, at 318.  
140. *Id.*  
141. Austin writes:  
   For example, take “voluntarily” and “involuntarily”: we may join the army or make a gift voluntarily, we may hiccup or make a small gesture involuntarily, and the more we consider further actions which we might naturally be said to do in either of these ways, the more circumscribed and unlike each other do the two classes become, until we even doubt whether there is *any* verb with which both adverbs are equally in place.  
   *Id.* at 323.  
142. George Fletcher has made the same point specifically in connection with eligi-
Throughout his essay, Austin refers to the excuses, but this part of his argument addresses the manner and circumstances of wrongdoing.\(^{143}\) In terms of the two components of culpability that I distinguished and defined at the outset, Austin addresses primarily issues of fault and nonfault, and not issues of eligibility and excuse. He notes that it is impossible to assess the significance of intentional states apart from the particular prohibition at issue,\(^{144}\) and that negligence as a kind of fault cannot be distinguished from negligence as a prohibition.\(^{145}\) Austin's argument tells us that there is an ineliminable connection between

---

\(^{143}\) Austin writes:

> Inattention, carelessness, errors of judgment, tactlessness, clumsiness, all these and other are ills (with attendant excuses) which affect one specific stage in the machinery of action, the *executive* stage, the stage where we *muff* it. But there are many other departments in the business too, each of which is to be traced and mapped through its cluster of appropriate verbs and adverbs.

*Austin, supra* note 17, at 324.

\(^{144}\) Austin writes:

> I sit in my chair, in the usual way—I am not in a daze or influenced by threats or the like; here, it will not do to say either that I sat in it intentionally or that I did not sit in it intentionally, nor yet that I sat in it automatically or from habit or what you will. It is bedtime, I am alone, I yawn: but I do not yawn involuntarily (or voluntarily!), nor yet deliberately. To yawn in any such peculiar way is just not to just yawn.

*Id.* at 322 (citation omitted).

\(^{145}\) Austin writes:

> We may plead that we trod on the snail inadvertently: but not on a baby—you ought to look where you are putting your great feet. Of course it *was* (really), if you like, inadvertence; but that word constitutes a plea, which is not going to be allowed, because of standards. And if you try it on, you will be subscribing to such dreadful standards that your last state will be worse than your first.

*Id.* at 324.
the primary norm—the conduct-governing directive to the population at large—and the applicable secondary norms, especially fault, that govern the enforcement of primary norms by legal authorities. In other words, fault is an aspect of wrongdoing, in the sense that the secondary norm of fault is coordinated with the primary norms that define criminal wrongdoing.\textsuperscript{146}

We can see this connection between primary and secondary norms in the Model Penal Code's construction of fault, in at least two different ways. First, the definitions of the "Kinds of Culpability" contained in section 2.02(2) make it clear that fault is inconceivable apart from a particular primary-normative context. Obviously, none of section 2.02(2)'s intentional states standing alone describes a kind of culpability. To act with a purpose with respect to a result or circumstance is to act with the objective of bringing it about.\textsuperscript{147} If I pat my baby's back with the objective of raising a burp, then I act with a purpose, but I do not act culpably. In order to act culpably, I must act with a purpose to bring about a result or circumstance which constitutes a crime. A purpose to kill is a culpable mental state, whereas a purpose to burp a baby is not a culpable mental state. The difference is that wrongdoing is involved in the former case, but not the latter.

Second, we cannot match just any mental state with just any circumstance, conduct, or result and thereby define a criminal offense. Section 2.02's four-level hierarchy of fault is perfectly reflected in a four-level hierarchy of offense degrees in only one instance: homicide.\textsuperscript{148} Other offenses are defined in terms of one or two of these fault levels, according to the nature of the offense. For example, fraud, or theft by deception, involves a purpose to deprive another of his property and conduct with a pur-

\textsuperscript{146} I will also refer to this connection as the primary-normative aspect of fault. For the sake of clarity, let me state that I do not mean to say that fault serves a primary-normative function. Fault is a secondary norm: whether it is taken to be only a necessary condition on punishment or is taken instead to be an affirmative reason for punishment, the question of fault is a question of the governance of legal authorities.

\textsuperscript{147} See Model Penal Code § 2.02(2)(a) (1985) (an act is purposeful when it is one's "conscious object to engage in conduct of that nature or to cause such a result").

\textsuperscript{148} See id. §§ 210.1-210.4.
pose to deceive.\textsuperscript{149} To substitute recklessness for purpose in this formulation would not define a different degree of fraud; it would not define a recognizable fraud at all.

To recognize the primary-normative aspect of fault—the dependence of fault's definition upon the particular wrongdoing at issue—undermines Hart's analysis of culpability. Hart contends that mens rea, fault, serves to distribute punishment, and to limit criminal liability in the same way that excusing conditions do; and he denies that proof of a particular mental state has anything to do with the justifying purpose of punishment. However, because fault is an aspect of wrongdoing, we cannot separate fault from the justification of punishment as a practice in the way that Hart supposes.\textsuperscript{150} If, as deterrence theory claims, the justification of punishment is a matter that pertains to the legislatively-defined offenses of a legal system, then the justification of punishment is also a matter that pertains to the fault that is conceptually and functionally inseparable from the material elements that constitute those offenses. Austin's thesis also implies that fault is an affirmative, justifying reason to punish, and not merely a necessary condition on punishment. If a violation of a legislatively-defined offense is an affirmative, justifying reason to punish, and not only a necessary condition for punishment, then the fault that is conceptually and functionally inseparable from this violation is also an affirmative, justifying reason to punish, and not only a necessary condition for punishment. In other words, all fault is genuine fault—an affirmative, justifying reason to punish, and not merely a side constraint on punishment.

Hart's arguments to the contrary in \textit{Punishment and Responsibility} seem perfectly sound, but this is attributable to the fact that Hart denies the very existence of fault in the criminal law, and addresses only excusing conditions. It is true that the excuses are distinct from the justification of punishment as a practice, that is, from the justification of the legislatively defined and enacted prohibitions that constitute the system of punishment. This is so because eligibility considerations are fundamentally

\textsuperscript{149} See \textit{id.} § 223.3.

\textsuperscript{150} See \textit{supra} notes 72-81 and accompanying text.
different from questions of fault. Austin's thesis plays out differently in connection with excusing conditions and fault, respectively.

It is important to note that, in cases in which punishment is authorized, we speak of the presence of fault and the absence of excusing conditions. Even if excusing conditions were strongly determined by the nature of the wrongdoing at issue, it would not follow that the absence of excusing conditions has anything to do with the justification of punishment. The absence of excusing conditions is only a feature of the ordinary, baseline context in which wrongdoing may or may not occur. Therefore, to accept Austin's thesis does not require us to say that eligibility and the excuses are implicated in the justification of punishment. Nor does it require us to say that eligibility or the absence of excusing conditions is an affirmative, justifying reason to punish—a proposition that contradicts existing law that we deem fundamental, because it implies that punishment is justified presump-
tively for each of us.151

In short, fault has a primary-normative aspect that fundamentally distinguishes fault from the absence of excusing conditions. Whereas eligibility considerations are merely necessary conditions for punishment, fault is an affirmative, justifying reason to punish in the same way that the criminal prohibition itself is an affirmative, justifying reason to punish. This theoretical difference explains the principal practical differences between fault and eligibility. Whereas the defendant has the burden of proof on excuses, the government has the burden of proof on fault because fault is an affirmative, justifying reason to punish in the same way that the criminal prohibition itself is an affirmative, justifying reason to punish. The failure to prove fault results in a failure to prove the offense because fault is an aspect of the wrongdoing that the other material elements of the offense describe.

To recognize the primary-normative aspect of fault allows us, pace Hart, to recognize genuine fault. Indeed, we are compelled to do so. We cannot escape the obligation to account for fault as an affirmative, justifying reason to punish, because fault is im-

151. See In re Winship, 397 U.S. 358, 361-65 (1970) (holding that criminal offenses must be proved beyond a reasonable doubt, even in juvenile cases).
licated in the justification of punishment as a practice—in the justification of the system of primary prohibitions—and not only in the adjudication of cases that constitute instances of the practice. The deterrence theorist’s construction of fault as a legal rule for the distribution of punishment that is unrelated to the moral justification of punishment refuses to meet this demand.\textsuperscript{152}

The foregoing analysis raises at least one further question. If fault is an affirmative, justifying reason to punish, in what way is this reason to punish different from the affirmative, justifying reason to punish that is constituted by the violation of the criminal prohibition itself? Put another way, why is fault a conceptually distinct reason to punish that cannot be incorporated into wrongdoing? The short answer to this question is that fault is an inference about the particular circumstances of the accused and about the particular manner of his wrongdoing—an inference made in the course of adjudication, that is necessary in order to bring the justification of the prohibition to bear on the individual case of punishment in a justifying way. I will expand on this answer below, in Part IV, but the full answer to this question can be given only after laying more groundwork.

\textbf{C. Two Explanatory Illustrations}

Before moving on, let me illustrate two of the ideas about fault that I have introduced and upon which much of what follows will turn. The ideas are the primary-normative aspect of fault and the existence of nonintentional fault. These illustrations make a point that is central to my argument. Both retributivist and consequentialist interpreters of fault chronically fail to appreciate the primary-normative aspect of fault, and consequently misconstrue the relationship between wrongdoing, intentionality, and fault. They infer from the frequent concurrence of wrongdoing, intentionality, and fault that an act can be

\textsuperscript{152} This may account for the inability of deterrence theory to vanquish retributivism at midcentury and for retributivism’s resurgence in the 1970s. See supra note 56 and accompanying text. Retributivism’s construction of fault as a proportion or balance between past and future that mandates punishment is an effort, at least, to account for fault as an affirmative, justifying reason to punish.
blameworthy and punishable only when it is done with some consciousness of harm or risk.\textsuperscript{153} This is simply false. The conclusion that should be drawn from the frequent concurrence of intentionality, fault, and wrongdoing is not that intentionality confers fault on wrongful conduct, but instead that fault is an inference from wrongful conduct that is sometimes, but not always, denoted by intentionality. The unreasonably mistaken rapist is, by hypothesis, unaware of his victim’s nonconsent or the risk of nonconsent, but this rapist is at fault: there exists, beyond his wrongdoing, an affirmative, justifying reason to punish him such that a failure to punish him would be an injustice.

The primary-normative aspect of fault makes it impossible to reduce fault to an intentional state. The Supreme Court effectively acknowledged this fact in \textit{Tison v. Arizona},\textsuperscript{154} when it recognized that culpability for capital murder cannot be encompassed in an intentional-state definition:

\begin{quote}
A narrow focus on the question of whether or not a given defendant “intended to kill,” however, is a highly unsatisfactory means of definitively distinguishing the most culpable and dangerous of murderers. . . . Some non-intentional murderers may be among the most dangerous and inhumane of all—the person who tortures another not caring whether the victim lives or dies, or the robber who shoots someone in the course of the robbery, utterly indifferent to the fact that the desire to rob may have the unintended consequence of killing the victim as well as taking the victim’s property. This reckless indifference to the value of human life may be every bit as shocking to the moral sense as an “intent to kill.”\textsuperscript{155}
\end{quote}

\textit{Tison} repudiates the notion that all nonintentional fault amounts to negligence, and that fault of this kind must be confined to minor offenses involving minor punishments.\textsuperscript{156} Not

\textsuperscript{154} 481 U.S. 137 (1987).
\textsuperscript{155} \textit{Id.} at 157. The \textit{Tison} court held that, under the Eighth Amendment, reckless disregard for human life can support a death sentence in a felony murder case. See \textit{Id.} at 148, 158.
\textsuperscript{156} In \textit{Tison}, two brothers were convicted of murder and sentenced to die because they helped their father break out of prison and, soon thereafter, their father killed
surprisingly, orthodox criminal law scholars reacted negatively to *Tison*.\(^{157}\)

The same pattern was evident in connection with another Supreme Court case that rejected the orthodox construction of fault. *Tison* recognized the constitutional validity of nonintentional fault under the Eighth Amendment.\(^{158}\) In a recent case on intoxication as a defense to a charge of noncapital murder, the Court recognized the validity of nonintentional fault under the Due Process Clause.

Under an intentionalist construction of fault, voluntary intoxication is relevant to disprove fault, because a sufficiently intoxicated person cannot attain certain mental states.\(^{159}\) For example, a person who has a blood-alcohol content of over three times the legal limit for driving is arguably unable to recognize any consequences of action or to behave with any fixed reference to consequences. Voluntary intoxication, therefore, should be admissible to disprove any level of fault that is premised on an intentional state. In other words, voluntary intoxication operates in the same way that an unreasonable mistake operates within the intentionalist construction of fault. For these reasons, the defen-
dant in *Montana v. Egelhoff*\(^{160}\) contended that he was entitled to offer proof of his severe intoxication to answer the allegation that he had killed his two victims purposefully or knowingly.\(^{161}\)

Not surprisingly, the traditional understanding of voluntary intoxication is quite different from the intentionalist view. Voluntary intoxication was not thought to negate fault at common law. Indeed, intoxication was thought to be further evidence of fault—fault that manifestly is premised on the accused's character, attitudes, and dispositions, and not on an intentional state that is present at the time of the criminal act.\(^{162}\) In *Egelhoff*, the Supreme Court turned back the defendant's due process challenge to Montana's adoption of the traditional rule on intoxication and affirmed the validity of the nonintentional conception of fault that underlies this rule.\(^{163}\) Incidentally, the Model Penal Code excludes proof of voluntary intoxication to disprove recklessness, and—by the intentionalist logic of section 2.02(3)—does so arbitrarily.\(^{164}\)

Presumably, Robinson and Darley reject the Court's recognition of nonintentional fault in connection with capital murder and voluntary intoxication. In both cases, however, Robinson and Darley would be at odds with a political reality that is grounded firmly in an intuition contrary to their expectations: serious crimes and severe punishments need not be premised on an intentional state regarding harm or wrongdoing, present at the time of the criminal act.

---

161. See id. at 40, 56 (upholding the murder conviction of a defendant with .36 blood alcohol content); see also United States v. Fleming, 739 F.2d 945, 947, 949 (4th Cir. 1984) (upholding the conviction for murder of a defendant with .315 blood alcohol content).
163. See *Egelhoff*, 518 U.S. at 56 (holding that a state can constitutionally choose to follow the traditional rule, even to the extent of excluding evidence of voluntary intoxication, if offered to negate a purpose to kill).
164. See MODEL PENAL CODE § 2.08 cmt. 1 (positing a "general equivalence" between recklessness in getting drunk and recklessness while drunk). The Code's postulate of a "general equivalence" is a non sequitur. A person who behaves recklessly while drunk might well have taken elaborate precautions against doing so before he began drinking.
Hart's analysis of the intoxication defense, at least, is also relatively clear. He would treat intoxication in the same way that he treats mistake: as a question of eligibility, not fault.\textsuperscript{165} That is, intoxication would be counted among the excusing conditions, with the caveat that when the intoxication is voluntary it probably will not be sufficiently reasonable to constitute an excuse. The objection that this would constitute a conviction for negligence would not trouble Hart because this objection rests on the premise—embraced under the intentionalist construction of fault and rejected by Hart—that a valid conviction requires the proof of intentional-state fault in association with each and every element of an offense.\textsuperscript{166} This does not mean, however, that Hart would have agreed with the State of Montana or with the Supreme Court in the analysis of Egelhoff. For Hart, the question of intoxication is not a question of fault at all because he denies the existence of genuine fault—fault as an affirmative, justifying reason to punish—and assimilates the remaining liability-limiting function of fault into the matter of eligibility.\textsuperscript{167} Hart supports nonintentional criminal \textit{liability}, but he does so on straightforward consequentialist grounds—not because he thinks there is such a thing as nonintentional \textit{fault}.\textsuperscript{168}

\textbf{D. Fault and Legal Obligation}

One feature of Hart's analysis of negligence that we have not yet examined provides an additional clue to the nature of fault. This point is of particular interest because it adumbrates Kahan's analysis of fault in his "new path" deterrence theory. As we will see, Hart did not carry the point far enough, and Kahan did not advance the point rigorously enough to succeed in the articulation of a new conception of fault.\textsuperscript{169}

As noted above, Hart argued against Bentham's belief that negligence should not be punished, because the notion that such punishment could not deter is a non sequitur.\textsuperscript{170} Hart also had a second, more important argument in favor of negligence liability.

\textsuperscript{165} See supra text accompanying notes 120-28.
\textsuperscript{166} See HART, supra note 85, at 114.
\textsuperscript{167} See id.
\textsuperscript{168} See id. at 132-35.
\textsuperscript{169} See infra notes 292-331 and accompanying text.
\textsuperscript{170} See supra notes 82-86 and accompanying text.
Bentham's argument supposes that the law works as a system of commands backed by threats. According to Bentham, criminal liability for negligence is absurd because the negligent actor would not recognize that he was subject to a legal command at the time of his action.171 Hart counters that the law is not a system of commands backed by threats, but is instead a system of obligations.172 We do not suppose that the legal system requires us always to advert to our legal responsibilities.173 Instead, we suppose that people abide by the law because they incorporate their legal obligations into the ordinary conduct of their lives.174 The duty of due care in negligence is one such obligation and its operation in the law, including the criminal law, is unproblematic.175

Hart stresses the value of the law's operating in this way in terms of classical liberalism and the consequentialist promotion of social welfare.176 Hart's rejection of fault as a reason to punish177 and his assimilation of the balance of fault into the liability-limiting operation of excusing conditions178 also are motivated by these midcentury reformist concerns.

However, if we view Hart's account of legal obligation, not from the perspective of the freedom that it allows, but from the perspective of the burden that it nevertheless imposes on indi-

171. See BENTHAM, supra note 68, at 174.
172. See HART, supra note 85, at 133-34.
173. Hart noted that people internalize the law's rules. See H.L.A. HART, THE CONCEPT OF LAW 140 (2d ed. 1994) ("Our rule-complying behaviour is often a direct response to the situation, unmediated by calculation in terms of the rules."); see also id. at 88 ("The fact that rules of obligation are generally supported by serious social pressure does not entail that to have an obligation under the rules is to experience feelings of compulsion or pressure.").
174. See HART, supra note 14, at 44.
175. See id. at 48.
176. Hart writes: If . . . we turn back to criminal law and its excusing conditions, we can regard their function as a mechanism for similarly maximizing within the framework of coercive criminal law the efficacy of the individual's informed and considered choice in determining the future and also his power to predict that future.
Id. at 46; see also id. at 48 (asserting that "the real satisfaction that a system of criminal law incorporating excusing conditions provides for individuals in maximizing the effect of their choices within the framework of coercive law").
177. See id. at 39-40.
178. See id. at 28-29.
individuals, then we can easily generate an account, not only of genuine fault, but of genuine nonintentional fault.

To say that the law permits us to plan and carry out our lives subject only to a minimal body of legal obligations is to imply that these legal obligations do govern the planning and carrying out of our lives. Furthermore, to say that this kind of legal governance does not require us constantly to advert to the law implies that the law does not govern us only in our conscious instrumental reasoning on every occasion of choice and action; rather, the law also governs us at the level of motivation. These two ideas complement one another. What we may come to desire and to pursue depends on the kind of life we lead, so that, in the governance of our life plans, the law also governs our motivations.

Under this conception of law, criminal wrongdoing, the violation of a primary prohibition, might reflect a failure to plan and to carry out a life that incorporates an appropriate set of motivations. This failure might be constitutive of the violation of law along with the wrongdoing. That is, we might consider this failure to assemble a proper set of ends to be an affirmative, justifying reason to impose the law's sanctions, which is distinct from and complementary to the wrongdoing. In other words, Hart's conception of law as a system of obligations supports a conception of genuine fault.

Furthermore, fault of this kind might be identifiable in some wrongdoing even if we cannot point to some intentional state that is present on the occasion of the wrongdoing. This fault would lie in the longer-term conduct of the accused, in the planning and carrying out of his life as a whole, as reflected in the particular manner and circumstances of his wrongdoing. In this case, we would have an instance of genuine nonintentional fault.

The foregoing conception of fault certainly is not attributable to Hart.179 This conception of fault is altogether foreign to the consequentialism that informs Hart's legal theory. Instead, it is a virtue ethics conception of fault. One of the distinguishing features of virtue ethics is that it describes governance at the level

---

179. Although, if it is a fair string of inferences, it calls his rejection of genuine fault into question.
of motivation.\textsuperscript{180} Even if one cannot attribute this conception of fault to Hart, however, it remains the case that his conception of law as a system of obligations brings deterrence theory to the border of a virtue ethics theory of punishment.

This point in Hart's deterrence theory is the point at which Kahan's "new path" deterrence theory begins. However, as Part III explains, the new path of deterrence dead-ends not far from here, because Kahan pursues this idea under the continuing influence of a particularly rudimentary consequentialism known as neoclassical economics. As Part IV shows, only an informed embrace of virtue ethics as the basis of a legal theory can carry us from Hart's sophisticated consequentialist theory of punishment onto a genuinely new path.

III. DETERRENCE ON A NEW PATH

Hart's second argument against Bentham's conception of deterrence draws on Hart's conception of legal obligation. Deterrence works, not only on our instrumental, means-ends reasoning, but also on our choices of ends and on our motivations. This is the line of argument that defines the "New Path of Deterrence." Remarkably, however, Hart is not the guide. The idea has been developed by law and economics scholars, under the heading of "preference shaping."\textsuperscript{181} Dan Kahan's recent work has pursued this new deterrence theory under the influence of Lawrence Lessig's "social meaning" theory of law and economics.\textsuperscript{182}

The old economic model of deterrence relied on the idea of rais-

\textsuperscript{180} See infra text accompanying notes 295-304.
\textsuperscript{181} See, e.g., Dau-Schmidt, supra note 3, at 25-30; Kahan, supra note 4, at 603-04; Parker, supra note 99, at 758-60.
\textsuperscript{182} See Kahan, supra note 22, at 351 n.7 (noting that, in framing his conception of deterrence, Kahan is following Lessig's lead); see also Kahan, supra note 1, at 2482-84 (describing social meaning as one of several subjects of a new deterrence scholarship). Lessig argues that people inhabit a social reality that is made up of text and context; that the meanings people perceive and the meanings they ascribe to their own actions constitute complex networks of motivation; and that law is a response to collective action problems that arise within these systems of meaning. See Lawrence Lessig, The New Chicago School, 27 J. LEGAL STUD. 661, 666-69 (1998) (describing a "second generation" of law and economics scholars that includes Kahan); Lawrence Lessig, The Regulation of Social Meaning, 62 U. CHI. L. REV. 943, 1044 (1995) [hereinafter Lessig, Social Meaning]; Lessig, supra note 23, at 2186.
ing the cost of crime. Kahan argues that we can deter crime more effectively if we pay attention to what criminal behavior means to those who engage in it and to those who are the victims of it. For example, if we wish students not to carry guns to school, then we should consider what a gun likely means to an adolescent: maturity, independence, strength, respect. Some of our punitive responses will reinforce these meanings and some will undermine them. If we wish to deter this activity effectively, then we should adopt the latter kind of response.

Given that Lessig’s social meaning theory significantly revises law and economics, one hopes that Kahan’s deterrence theory will offer a significant revision of the deterrence theory’s conception of fault. Specifically, one hopes that Kahan’s revision will recognize fault as an affirmative, justifying reason to punish, acknowledge the primary-normative aspect of fault, and accommodate instances of nonintentional fault. These hopes are realized to some extent. At least, Kahan’s recent work begins to grapple with the problem of fault.

Unfortunately, Kahan fails to explain the key features of fault. As a result, he fails to explain the way in which the criminal law truly governs our motivations. Kahan frames deterrence as a question of practical reasoning beyond instrumental reasoning, as Hart’s arguments suggest one might, and Kahan recognizes that practical reasoning does not invariably involve the promotion of good consequences. However, the inveterate consequentialism of economic theory is a drag on these insights. Kahan disavows the aim of presenting a comprehensive theory of punishment, but it is clear that the new path of deterrence will not lead to a coherent theory of punishment in any event.

184. See Kahan, supra note 4, at 597 (citing Lessig’s social meaning theory in support of an “expressive” theory of punishment); see also Kahan, supra note 22, at 365 (arguing for descriptive and prescriptive applications of Lessig’s social meaning theory to criminal law); Dan M. Kahan, Social Meaning and the Economic Analysis of Crime, 27 J. LEGAL STUD. 609, 610 (1998) [hereinafter Kahan, Social Meaning] (arguing that the criminal law promotes approved meanings and suppresses meanings that are disliked or feared).
185. See Kahan, supra note 22, at 363-65.
A. Kahan on Mistake and Fault

Kahan's conception of fault can be drawn out of his analysis of mistake. We do not excuse crimes that are committed under a mistake of law, Kahan argues, because we want to deter "loopholing." That is, the letter of the law is rigid, and legislators cannot stay ahead of creative efforts to exploit gaps in the law. Our refusal to allow the mistake of law defense is one way in which we plug these gaps. The doctrine is an "injunction to do what's right rather than what one thinks is legal." This analysis does not apply to crimes that do not correspond to moral offenses. If a crime is merely *malum prohibitum*, a mistake of law will excuse. For example, because there is no background moral prohibition in tax and licensing cases, we are not concerned with "loopholing."

Prompted by Daniel Yeager, Kahan has acknowledged that his analysis applies not only to mistakes of law, but also to mistakes of fact. Kahan writes:

"Courts confine both types of mistake defenses to actors who have internalized community moral norms: if the defendant would be acting consistently with those norms were circumstances—factual and legal—as she supposed them to be, she gets a defense; if not, not. Both ignorance of fact and ignorance of law, then, are indeed excuses only for the virtuous."

The conception of fault that is implicit in this account of mistake seems promising. Kahan's conception of mistake acknowledges the primary-normative aspect of fault and recognizes fault as an affirmative, justifying reason to punish. If we punish in cases of mistake because the accused has violated a background morality, then fault is not only a matter of the distribution of punishment according to liability-limiting secondary norms. Kahan's conception of mistake also acknowledges genuine nonintentional

186. Kahan, supra note 2, at 138.
187. See id. at 139-40.
188. Id. at 141.
189. See id. at 138, 148-50.
fault. Kahan's argument implies that the unreasonably mistaken rapist would and should be convicted of rape, regardless of his having no mental state regarding nonconsent, because he has violated a background morality.

As the foregoing analysis of mistake in rape suggests, and as Professor Yeager has pointed out, however, Kahan's conception of mistake raises a problem. Kahan leaves the background morality unstated and expressly has declined Yeager's invitation to articulate it. Furthermore, at least two other, related problems lurk in Kahan's argument. Kahan fails to articulate not only the nature of his background morality, but also a jurisprudence that would explain why any given morality should be enforced legally. Also, Kahan does not explain what he means by "internalized community moral norms." We are familiar with the phenomenon of internalized norms. As noted above, Hart suggests the possibility of internalizing legal norms in an argument that suggests a virtue ethics legal theory. Kahan's reference to "the virtuous" in this connection strongly suggests a virtue-ethics approach to this set of issues. But even though Kahan makes one reference to a virtue-ethics theory of punishment, he does not pursue it with any rigor.

Instead, Kahan pursues his explanation of fault within the confines of consequentialism and draws on the cognate discipline of law and economics. Kahan argues that mistake will be rec-

192. See id. at 2127.

193. In terms of Kahan's argument, the unreasonably mistaken rapist will not have a defense of mistake because, even if the circumstances had been as he supposed, he could not reasonably infer consent from those circumstances. His nonconsensual sexual intercourse would have violated a background morality. Alternatively, even if consent had been among the circumstances as he supposed them to be, his consensual sexual intercourse would have violated a different part of the background morality. See Anne M. Coughlin, Sex and Guilt, 84 VA. L. REV. 1, 6 (1998) (arguing that rape laws are artifacts of a background moral prohibition on fornication instead of the prohibition on assault).

194. See Yeager, supra note 190, at 2116.

195. See id. at 2116-20 (offering an invitation); Kahan, supra note 191, at 2127 (declining the invitation).

196. Kahan, supra note 191, at 2125.

197. See supra text accompanying notes 172-80.

198. See Kahan, supra note 2, at 129 n.9, (referring the reader to Huigens, supra note 6, and Kahan & Nussbaum, supra note 6).

199. See Kahan, supra note 1, at 2478-79.
ognized as a defense when the mistake occurs at the edges of moral consensus. In these cases, liability for mistaken offenses would overdeter through a chilling effect on activities that otherwise would generate "licit utility." Kahan continues:

However, when the fact in question does not sit on the moral boundary line, but is actually located within the interior of what's immoral, then the chilling effect of strict liability is good because any utility from the forgone activity would have been illicit: if a nervous mugger refrains from assaulting a private citizen because he fears that the person could be a law-enforcement official, all the better from society's point of view.

This argument prompts two objections. First, the distinction between "licit" and "illicit" utility is mysterious. Only one kind of utility exists and there is only more and less of it; to say otherwise is to jettison the economist's assumptions that value is monistic and that the unique rational response to value is to promote it. Certainly, one can reject these assumptions, but then one would not be talking about utility in any sense that is recognizable from an economic perspective. Second, Kahan suggests that in a case such as that of an assault on a police officer, the chilling effect of the mistake doctrine results in overdeterrence that is somehow good. This too is mysterious from an economic perspective. Any overdeterrence is suboptimal by definition, and suboptimality is, by definition, a bad rather than a good result.

As it happens, Kahan's terminology has an impeccable pedigree in the economic analysis of the criminal law. George Stigler coined the phrase "illicit utility" to describe the social value of criminal activity in an article published in 1970, in response to Gary Becker's groundbreaking article of 1968. The phrase "illicit utility" means only that one can plausibly set the value of the murder to the murderer and of the rape to the rapist at

---

200. See Kahan, Social Meaning, supra note 184, at 610.
201. Kahan, supra note 191, at 2126.
202. Id. at 2126-27.
203. See infra note 222 and accompanying text.
204. See Stigler, supra note 101, at 527 (citing Becker, supra note 183).
naught in the social welfare calculus.\textsuperscript{205} However, this reasonable interpretation of the phrase should not cause us to overlook some important facts. First, Becker, not Stigler, is the consistent economist: the murder does have value to the murderer, and to exclude this value from one's social welfare calculations can be done only by stipulation.\textsuperscript{206} Second, the plausibility of this stipulation depends on traditional, noneconomic categories of wrongdoing and fault.\textsuperscript{207} Third, economic analysis simply cannot generate, from its own premises, any adequate version of these criminal categories.\textsuperscript{208} In sum, to speak of "illicit utility" may be a reasonable shorthand for a plausible valuation of zero, but the fact that the phrase is an oxymoron should serve to remind us that the valuation of zero itself involves an appeal to concepts that are ill-fitted to the basic tenets of economic analysis. Stigler's hybrid notion, which, in one form or another, appears in most economic analyses of the criminal law,\textsuperscript{209} ultimately is untenable without some satisfactory explanation of its reliance on the traditional values and categories of the criminal law.

Even though Kahan does not say so,\textsuperscript{210} I suspect that Lawrence Lessig's theory of "the regulation of social meaning" is

\textsuperscript{205} See id.

\textsuperscript{206} See Parker, supra note 99, at 759 (arguing that the adverse welfare effects of overly punitive penalties cannot be so easily disposed of).

\textsuperscript{207} See id. at 760 ("Like the standard economic model, [the illicit utility] variations take the definition of crime as exogenous to the enforcement problem."); see also David D. Haddock et al., An Ordinary Economic Rationale for Extraordinary Legal Sanctions, 78 CAL. L. REV. 1, 12-13 (1990) (arguing against the "illicit utility" concept on the ground that "[w]idespread attitudes that particular behavior is 'immoral' or that certain utility is 'socially illicit' are endogenous phenomena to be explained, not exogenous phenomena that do the explaining").

\textsuperscript{208} See Alvin K. Klevorick, On the Economic Theory of Crime, in CRIMINAL JUSTICE, supra note 10, at 289, 303 (arguing that "an explanation of the criminal category—even an explanation stated in economic terms—requires answers to questions and elaboration of concepts that economic analysis is not particularly well-suited to provide").

\textsuperscript{209} See, e.g., Shavell, supra note 99, at 1234 (arguing that a distinction between social and private benefits "gives the analyst greater freedom to describe a society's values," including the valuation of crime at zero); see also Parker, supra note 99, at 759 (listing works by Posner, Shavell, Cooter, Klevorick, and McChesney that make this assumption).

\textsuperscript{210} Kahan declines consistently to engage in "deep theorizing," Kahan, supra note 4, at 597, and his scholarly agenda is expressly political, rather than philosophical. See Kahan, supra note 1, at 2478-79. As a result, it is difficult to tell how Kahan would account for his use of the term "illicit utility," or for the role of Lessig's social meaning theory in Kahan's analysis of mistake and, by extension, of fault.
meant to play this role in Kahan's "new path" deterrence theory. 211 That is, Kahan hopes to use Lessig's social meaning theory to connect the economic analysis of law to the background morality of the criminal law.

Lessig's notion of "social meaning" is, like Stigler's notion of "illicit utility," a hybrid of economic and noneconomic ideas. Lessig recognizes certain features of value and practical reasoning that both consequentialist ethics and neoclassical economics have long neglected, and he packages these noneconomic valuations, motivations, actions, and norms for economic analysis, under the label of "social meaning." 212 For example, we seem to value children in ways that are different from the ways in which we value commodities such as whipping cream or boots. We ascribe intrinsic value to children; that is, we do not value them only for the sake of their use in satisfying our own preferences. We also ascribe incommensurable value to children; that is, we do not treat them as fungible goods. Neither intrinsic value nor incommensurable value is recognized by the economist's conception of value as revealed preference. 213 Lessig argues, in effect, that these values can be recognized in economic analysis if we suppose that people have preferences for the expression of these values, and that they act for the optimal satisfaction of these preferences for "social meaning," just as they do for any other preferences. 214

Kahan argues for an "expressive theory" of deterrence along these lines.

211. Kahan does say this much:

The phenomena of social influence and social meaning matter for deterrence. The decisions of individuals to commit crimes are influenced by their perception of others' beliefs and intentions; the law shapes information about what those beliefs and intentions are. It follows that a community that wants to deter crime should concern itself not just with the effect of particular policies on the price of crime but with the statements that those policies make (and enable others to make) about the public's attitudes toward criminal behavior.

Kahan, supra note 22, at 351.

212. See Lessig, Social Meaning, supra note 182, at 951-52.

213. See infra text accompanying notes 269-79.

214. See Lessig, Social Meaning, supra note 182, at 1001-04 (arguing that the ritualistic mourning of the Hindu widow can be modeled as the optimization of the satisfaction of a preference for the expression of Hindu values).
The expressive theory can also be used to inform deterrence. One way in which it might do so is by supplying a consequentialist theory of value. Without a theory for identifying which outcomes are socially disvalued and how much, it is impossible to know what to deter or how to allocate limited punishment resources among different forms of wrongdoing. Again, one could attempt to specify a consequentialist theory of value—perhaps wealth-maximization—that is indifferent to expressive sensibilities. Or one could overtly draw on these sensibilities to identify preferred outcomes. To return to an earlier example, one might conclude that a white man who kills an African-American out of racial hatred should be punished more severely than a woman who kills the abuser of her child in anger, even if equal punishment would maximize social wealth; when expressive considerations are taken into account, racist killings are deemed to harm society more than are impassioned killings of child molesters.\(^{215}\)

In other words, if a standard economic conception of value such as wealth or preference satisfaction does not capture the intrinsic value of an African-American life or a child’s life, then we can capture these values for economic analysis by treating our beliefs about the intrinsic value of such lives as preferences that are to be optimized within the criminal justice system.

In connection with mistake, Kahan argues in a similar vein that “[b]y denying a mistake of law defense, the law is saying, contra Holmes, that if a citizen suspects the law fails to prohibit some species of immoral conduct, the only certain way to avoid criminal punishment is to be a good person rather than a bad one.”\(^{216}\) This argument is Lessig-like in that it claims the law creates preferences for morality-abiding behavior that we optimize along with our other preferences.\(^{217}\)

If Kahan does hope to use Lessig’s social meaning theory to connect the economic analysis of the criminal law to the background morality of the criminal law, then Kahan’s “new path” deterrence theory is at a dead end. Lessig’s hybrid theory is

\(^{215}\) Kahan, supra note 4, at 602-03 (citations omitted).
\(^{216}\) Kahan, supra note 2, at 129.
\(^{217}\) See Kahan, supra note 4, at 603-04 (arguing that “law also moralizes by shaping relevant 'belief-dependent' preferences”).
simply untenable. One indication of this fact is that Kahan's argument about mistake is much less an anti-Holmesian account of the criminal law than it is a super-Holmesian account.\textsuperscript{215} It contends only that Holmes's bad man can be made to engage in strategically moral behavior. A genuine anti-Holmesian account of the criminal law would distinguish this kind of morality-promotion from virtue. Genuinely virtuous behavior springs from proper motives. Strategically virtuous behavior, of the kind that Kahan describes, is a contradiction in terms.\textsuperscript{219} This distinction is a critical one, as we will see, because the criminal law does have a role to play in the acquisition of proper motivations.

B. Ad Hoc Nonconsequentialism

Neither Kahan nor Lessig seems to realize that, when he recognizes the existence of nonconsequential value and practical reasoning, he undermines the consequentialist premises of economics. Broadly speaking, the "social meaning" half of Lessig's theory recognizes all the essentials of a pluralistic theory of value, and at least a modest cognitivism in the theory of motivation; but the law and economics half of Lessig's theory continues to insist that value is monistic, and that motivation is a matter of noncognitive desire. These are fundamentally incompatible philosophical commitments, and Lessig's attempt to combine them is simply incoherent. This incoherence infects Kahan's account of fault.

Neoclassical economics is a form of consequentialism that is premised on specific conceptions of value and practical reasoning

\textsuperscript{218} See Kahan, supra note 2, at 129.
\textsuperscript{219} See ARISTOTLE, supra note 25, at 45 (describing virtue as a fixed disposition toward the good); see also L.A. Kosman, Being Properly Affected: Virtues and Feelings in Aristotle's Ethics, in ESSAYS ON ARISTOTLE'S ETHICS 103, 108-09 (Amelie Oksenberg Rorty ed., 1980) [hereinafter ESSAYS] (describing virtue as action from proper motives, construed as emotional states). This view is not limited to classical ethics; it is also a feature of Christianity. Christian salvation requires a changed heart and right conduct from proper motives. See JONATHAN EDWARDS, Sinners in the Hands of an Angry God, in JONATHAN EDWARDS: REPRESENTATIVE SELECTIONS 155, 161 (Clarence H. Faust & Thomas H. Johnson eds., rev. ed. 1962) ("[I]t is plain and manifest, that whatever pains a natural man takes in religion, whatever prayers he makes, till he believes in Christ, God is under no manner of obligation to keep him a moment from eternal destruction.").
that have dominated Anglo-American social thought for centuries. The economist's central assumptions are that value is a function of personal preferences; that value is commensurable in terms of utility, social welfare, or wealth; that practical reasoning consists of instrumental reasoning for the optimization of value; and that social choices can be made according to the optimization of value so conceived. These assumptions are underwritten by noncognitivist conceptions of value and practical reasoning. The basic economic concept of a preference has its roots in Hobbes's view that value can be reduced to pleasure and pain. The basic economic concept of preference satisfaction is rooted in the Humean concept of motivation as desire.

---

221. See, e.g., Leonard J. Savage, The Foundations of Statistics 69 (2d ed. 1972) ("A function $U$ that thus arithmetizes the relation of preference among acts will be called a utility. It will be shown that the multiplicity of utilities is not complicated, every utility being simply related to every other.").
222. See, e.g., Gary S. Becker, The Economic Approach to Behavior 153 ("Everyone more or less agrees that rational behavior simply implies consistent maximization of a well-ordered function, such as a utility or profit function."); Savage, supra note 221, at 96-104 (defending the notion of maximizing expected utility as both a descriptive and normative account of rational choice); id. at 158-59 ("[T]he minimax theory is predicated on the idea that the consequences of those acts with which it deals are measured numerically by a quantity of the expected value of which the person wishes to have as large as possible."). A related feature of the economist's conception of practical reasoning is its prospective orientation. See Posner, supra note 101, at 7-8.
224. See Hobbes, supra note 77, at 39-40 ("But whatsoever is the object of any mans Appetite or Desire; that is it, which he for his part calleth Good: And the object of his Hate, and Aversion, Evill; . . . "). Becker acknowledges the roots of economics in this hedonic conception of value. See Becker, supra note 222, at 8 (describing the basis of economic theory in "the pleasure-pain calculus"); see also Martin Hollis & Robert Sugden, Rationality in Action, 102 Mind 1, 2-3 (describing Hobbes's conception of value as among the origins of rational choice theory).
225. On the Humean view, my choice of means to the satisfaction of my desire can be affected by reasoning, but this is the first phase of an action at which reason and reasons can operate. See David Hume, A Treatise of Human Nature 414 (L.A. Selby-Bigge & P.H. Nidditch eds., 2d ed., Oxford Univ. Press 1978) (1739-40) ("But 'tis evident in this case, that the impulse arises not from reason, but is only directed by it."). A second formulation is quoted more frequently: "Reason is, and ought only to be the slave of the passions, and can never pretend to any other office than to serve
people's valuations and motivations are set in a uniformly arational way, at the level of brute pleasure and pain, then one can assume that people will strategically optimize their preference satisfaction and that their revealed preferences represent commensurable value.226

In the discussion that follows, I will identify six features of value and practical reasoning that Kahan, following Lessig, recognizes in his "new path" deterrence theory: the cognitive dimension of valuation and motivation; rationally expressive acts; intrinsic value; local systems of value; the incommensurability of value; and deliberation on ends. Individually and collectively, these six features of value and practical reasoning undermine the noncognitivist and consequentialist premises of economics.

1. Preferences, Valuation, and Motivation

Kahan follows Lessig when he argues that people ascribe meaning to their actions and that their preferences can be shaped by laws that address these meanings.227 For example, Kahan argues that anti-gang legislation ought to take into account the social meanings associated with gangs.228 Gang membership denotes a kind of strength that is very attractive to adolescents: strength in numbers.229 In communities in which gang activity is prevalent and highly visible, the power that comes with belonging to the group may appear to be a necessity to young people, in spite of the fact that, as individual members of the community, they fully appreciate and abhor the violent and

and obey them." Id. at 415.

226. Modern economic theory disavows any reliance on these psychological premises. See Hollis & Sugden, supra note 224, at 2-7. Preference is considered the basis of the economist's theory of value because preferences are revealed in completed choices. This relieves the economist of the burden of making intersubjective comparisons of value according to pleasure or desire. See SAMUELSON, supra note 220, at 90-92. This modification does not surrender the noncognitivist commitments of economic theory, however. Completed choices reveal only primitive likes and dislikes. These attitudes are not the sole source of value; they are only the part of value that Hobbes described. See infra notes 248-52 and accompanying text.

227. See Kahan, supra note 22, at 351 n.7 (noting that, in framing his conception of deterrence, Kahan is following Lessig's lead); see also Kahan, supra note 4, at 603-04 (describing the role of preference shaping in deterrence theory).

228. See Kahan, supra note 22, at 374.

229. See id. at 375.
destructive nature of gangs. The aggressive demeanor of gang members is self-perpetuating in the same way: in a place where such a demeanor is common, to adopt an aggressive demeanor oneself becomes a defensive or preemptive necessity. The net effect is a "system of shared misunderstanding" in which the life of a gangster—including the commission of crime—is preferred. Highly publicized "crackdowns" on gang activities fail as policy because they fail to address this dimension of the gang phenomenon. Kahan contends that gang loitering ordinances, in contrast, are preference-shaping criminal laws, and are effective in deterring crime for precisely this reason.

When Kahan argues that preferences can be altered in these deliberate ways, by means of articulate appeals to one's identity and attitudes, he effectively repudiates the Humean conception of value and motivation. In doing so, he repudiates the economist's conception of value as preference that is revealed in completed choice. Hume argued that reason has no role in the shaping of motivation; that motivation is a function of non-cognitive desire. The economist maintains this view, sometimes in these explicitly psychological terms, but always implicitly in his treatment of preferences as observable data which must remain fixed for purposes of economic analysis. Kahan does not argue merely that a person's desires can be subliminally conditioned, which would be consistent with non-cognitivism. He

230. See id. at 374.
231. See id. at 375.
232. Id. at 374.
233. See id. at 374-75.
234. See id. at 375-76.
235. See id. at 376.
236. See HUME, supra note 225, at 414-15.
237. Becker writes:

   The economic approach to human behavior is not new, even outside the market sector. Adam Smith often (but not always!) used this approach to understand political behavior. Jeremy Bentham was explicit about his belief that the pleasure-pain calculus is applicable to all human behavior: "Nature has placed mankind under the governance of two sovereign masters, pain and pleasure. It is for them alone to point out what we ought to do, as well as to determine what we shall do. . . . They govern us in all we do, in all we say, in all we think." BECKER, supra note 222, at 8 (quoting BENTHAM, supra note 68, at 1).
238. See id. at 5.
argues instead that motivations can be altered by an articulable appeal to a person's self-conception, and to the logical implications of a given action for this self-conception.\textsuperscript{239} This is to recognize, contra Hume and the economist, that valuation and motivation have a cognitive dimension. Perhaps there is a kind of consequentialism that can sustain a repudiation of strict noncognitivism. Kahan's difficulty is that economic theory is not this kind of consequentialism. Its central concept of a preference is too heavily dependent on the non-cognitivist conceptions of value as pleasure and of motivation as desire.

In order to see this shortcoming in law and economics more clearly, consider a more sophisticated neo-Humean theory of value and motivation. Like Hume,\textsuperscript{240} Gerald Gaus argues that valuation is an affective—that is, an emotional—response to the objects of one's attention.\textsuperscript{241} Gaus argues that "intrinsic valuing are emotional dispositions."\textsuperscript{242} But Gaus's theory employs a more sophisticated conception of emotion than any Hume had in

\begin{footnotesize}
\begin{enumerate}
\item[239.] See Kahan, supra note 22, at 349-52.
\item[240.] Hume shared Hobbes's conception of value: "'Tis from the prospect of pain or pleasure that aversion or propensity arises towards any object . . . ." Hume, supra note 225, at 414.
\item[241.] See GERALD F. GAUS, VALUE AND JUSTIFICATION: THE FOUNDATIONS OF LIBERAL THEORY 49-64 (1990). If Gaus's neo-Humean theory is not to one's liking, one can choose other theories. Like Gaus, Michael Smith argues that belief and desire are connected more intimately than Hume supposed, but for reasons different from those that Gaus gives. Smith sees the cognitive dimension of motivation in the conscious acquisition, and consistent ordering, of one's dispositions. See Michael Smith, The Moral Problem 92-129 (1995) (setting forth a neo-Humean theory of motivation); see also Philip Pettit & Michael Smith, Practical Unreason, 102 Mind 53 (1993) (describing practical reasoning in terms of the consonance of deliberative reasons and desire). Bernard Williams sets out a similar conception of motivation. See Bernard Williams, Internal and External Reasons, in Moral Luck: Philosophical Papers 1973-1980, at 101, 102-04 (1981) (describing deliberations over and adjustments to one's "motivational set" as part of a neo-Humean conception of motivation). Smith and Williams part company over whether these deliberations provide relative or nonrelative normative reasons. Smith departs ultimately from Hume when he concludes that the reinterpretation of desire as the deliberate acquisition of dispositions results in a theory of nonrelative rather than relative normative reasons. See Smith, supra, at 165-75. In contrast, Williams argues that one's new dispositions necessarily remain relative to previous dispositions. In this way, Williams maintains an essentially Humean view. See Williams, supra, at 105-11.
\item[242.] GAUS, supra note 241, at 25.
\end{enumerate}
\end{footnotesize}
mind—one which is well supported by empirical study. Gaus's "Affective-Cognitive Theory" of the emotions maintains that emotions are both affective states (that is, feelings) and intentional states (that is, mental states directed toward an object). This implies that emotions can be rationally criticized because they have proper objects and give one reasons for appropriate attitudes and actions. Moreover, if valuations consist of these affective-cognitive responses, then value can be meaningfully deliberated upon.

Similarly, one need not resort to cognitivism per se—the position that motivation can follow from belief, without an affective component—in order to appreciate the cognitive dimension of motivation. A sophisticated neo-Humean theory will do. Valuations conceived of in Gaus's terms, as affective-cognitive responses, directly imply reasons for action. As Gaus puts it: "[I]t seems completely implausible to say that one who is experiencing fear, and believes that the fear is grounded in the threatening characteristics of a tiger and that running away will remove the danger, has no reason whatever to run."

The difficulty for Lessig and Kahan is that, whereas they clearly want to embrace a neo-Humean theory of value and motivation such as Gaus's, their commitment to law and economics precludes this. Economics is a Humean enterprise, but it cannot keep up with its philosophical cohort. Economic theory relies on a revealed preference theory of value, according to which all value is to be inferred from rational agents' completed

243. See id. at 49-64.
244. See id. at 49-56.
245. See id. at 64-79. For example, suppose the following is a true report of my internal state: I am scared that that baby over there is about to spring up and kill me, but even though I am in fear for my life I am motivated to continue to sit where I am rather than to get up and run away. Something (indeed, everything) is wrong here. First, the object of my emotion is inappropriate; a baby will not inspire fear of death in one who is not crazy in some way. Second, assuming that I am genuinely, though crazily, in fear, my motivation and action are wrong: fear gives one a reason to flee, not to stay.
247. GAUS, supra note 241, at 136. In this scheme of affect, belief, and action, the archaic philosophical concept of desire as a distinct motivating force is simply unnecessary. See id. at 164-65.
choices. But neo-Humean theories of value undercut the revealed preference conception of value.

A completed choice is either pro or con. This simple binary signal reveals a general affective response: like or dislike. Furthermore, a completed choice necessarily reveals only this much of the agent's affective state. This is a problem for the revealed preference theory of value. Many valuations, like the emotions on which they are premised, are far more complex and subtle than like and dislike. If we recognize the cognitive dimension of emotion, then we can see why this is so. Emotions fall along a spectrum of specificity that corresponds to the specificity of their proper objects and manifestations.\(^{248}\) Some emotions are highly specific: emotions such as indignation, pity, and resentment are associated with quite distinctive occasions and behaviors. Other emotions, notably liking and disliking, are more general. The valuations that are premised on these emotions fall along a parallel spectrum,\(^{249}\) and in this spectrum of valuations we can see a reason to reject the revealed preference theory of value. Only the specific behavior in which an emotion is manifested can reveal the detail of the emotion; and the same is true of any valuation that the emotion constitutes. The economist's conception of value as preference revealed in completed choice necessarily misses this detail.

This loss of detail in the representation of value undermines the economist's preference theory of value completely, not partially.\(^{250}\) It is tempting to suppose that the loss of detail is benign—that the core or essence of value nevertheless is indicated or represented in revealed preferences. But this is a mistake. To

\(^{248}\) See id. at 65 ("In relation to the appropriateness of their grounding beliefs, then, we would do best to envisage emotions on a spectrum, ranging from the very specific to the open-ended.").

\(^{249}\) This point corresponds to a familiar distinction in philosophical ethics between "thick" evaluative concepts—indignation, pity, resentment—and "thin" evaluative concepts—for example, "good," "bad," etc. See Elizabeth Anderson, Value in Ethics and Economics 98-99 (1993) (citing, inter alia, Philippa Foot, Moral Arguments, in Virtues and Vices and Other Essays in Moral Philosophy 96 (1978); David Wiggins, A Sensible Subjectivism?, in Needs, Values, Truth: Essays in the Philosophy of Value 185 (1987); Bernard Williams, Ethics and the Limits of Philosophy 140 (1985); John McDowell, Virtue and Reason, 62 Monist 331 (1979)).

\(^{250}\) See Gaus, supra note 241, at 84-112 (critically examining theories of valuation premised on desire, preference, pleasure, and satisfaction).
fail to capture the detail of valuation is not to capture part, albeit not the whole, of all values. It is, instead, to fail to recognize some values at all: those that are premised upon emotions that have particular occasions and distinctive behavioral manifestations.\textsuperscript{251} Revealed preference simply is not an adequate conception of value because it is blind to so many values.\textsuperscript{252} Kahan implicitly recognizes this inadequacy—and effectively repudiates economics—when he recognizes the cognitive dimension of value and motivation under the heading of “social meaning.”

2. Economics and the Expression of Value

Kahan argues that punishment has an expressive dimension that we ignore at our peril.\textsuperscript{253} In doing so, he repudiates another central element of the economist’s conception of value and practical reasoning: the assumption that the unique rational response to value is action for the sake of the optimization of value. This point is related to the preceding one: valuations that are premised on particular occasions and behaviors can only be manifested on these occasions and by these behaviors. That is, some values exist only in their expression; and it is therefore impossible to suppose, as the economist does, that the rational response to value is always to optimize it.

Kahan acknowledges that some values exist only in their expression and that the rationality of some actions consists in their being expressive. For example, Kahan argues that the standard economic account of punishment overlooks the significance of alternative sanctions. He writes:

This account is defective because it ignores what different forms of affliction mean. Punishment is not just a way to

\textsuperscript{251} See infra note 279 (considering the economist’s insistence that admiration is a matter of pure taste that can and ought to play no role in genuine practical reasoning).

\textsuperscript{252} One might argue that the general affective responses of like and dislike can serve as general categories under which more specific emotions and valuations can be grouped, and that the pro and con attitudes that completed choices reveal are therefore a sufficient foundation for the preference theory of value. This is to concede, however, that preferences do not constitute, reveal, or correspond to values themselves, but only to generalizations about values. At best, we might employ the preference theory in conjunction with a genuine theory of value.

\textsuperscript{253} See Kahan, supra note 4, at 594-605.
make offenders suffer; it is a special social convention that signifies moral condemnation. Not all modes of imposing suffering express condemnation or express it in the same way... 254

Kahan portrays his argument for alternative sanctions as a special instance of a general account of expressive rationality, which he illustrates with this argument and example:

Part of being rational consists in selecting actions that, against the background of social norms, express meanings appropriate to our purposes and goals. Along some dimension, for example, five thousand dollars might be equivalent in value to everything I would do with and for a friend during a certain period of time. But if my goal is to be her friend, then giving her the money and sharing my time with her are not interchangeable; giving money in lieu of time fails to convey the respect and affection that being a good friend requires. 255

If Kahan believes that expressive rationality of this kind can be incorporated into the consequentialist framework of economic analysis, then he has misunderstood expressive rationality. 256

Christine Swanton frames the issue in a way that Kahan would appear to endorse:

It is a common mistake, as Michael Stocker points out, to believe that action done for a reason, i.e., motivated action, is always action done for the sake of some end. The idea that the end must be future-directed, as in consequentialist models of rationality, is a relatively extreme species of this mistake. But some motivations are not goal-directed at all, but are "expressive," as in "I did it out of friendship." 257

254. Id. at 593.
255. Id. at 597.
256. Kahan cites Elizabeth Anderson's Value in Ethics and Economics as authority for a general account of expressive rationality. See id. at 597 n.21. Anderson makes it quite clear, however, that expressive rationality and the consequentialist conception of rationality are fundamentally different and incompatible conceptions of practical rationality. See ANDERSON, supra note 249, at 32 ("Consequentialist and expressive theories of rationality pose sharply contrasting ways of thinking about value and action.").
In other words, to recognize the practical rationality of expressive actions is to repudiate the consequentialist conception of practical rationality, if only because expressive actions are retrospectively as well as prospectively oriented. As the work of Swanton, Stocker, and others makes clear, to act expressively is precisely not to act in a way that will optimize the satisfaction of one's preferences. Expressive action is not directed toward the production of consequences at all, which makes expressive actions arational, at best, from an economic perspective. Kahan cannot have it both ways. If he recognizes expression as a rational response to value, then he repudiates a central premise of economics: the idea that a rational response to value always amounts to action for the optimization of preference satisfaction.

In fact, only the rhetoric of social meaning theory acknowledges expressive actions; the analysis is thoroughly consequentialist. For example, Lessig purports to give the expressive dimensions of Hindu widowhood their due when he argues that the Hindu widow who wears rough clothing and confines herself to her home after the death of her husband is optimizing a preference for social standing that is conditioned on her making these traditional expressions of grief. But when he packages the Hindu widow's expressions of grief as preferences, and then introduces them into an unmodified consequentialist framework, Lessig does not really recognize expressive rationality at all. To recognize the expression of value in this case requires one to acknowl-

---


259. Justin Oakley argues that the feature of virtue ethics that finally distinguishes it from consequentialism is virtue ethics' rejection of maximization as a criterion of rightness. See Oakley, supra note 26, at 143, 149-51.

260. See SMITH, supra note 241, at 130 ("The distinctive Humean view of normative reasons, then, is that the rational thing for an agent to do is simply to act so as maximally to satisfy her desires, whatever the content of those desires.").

261. See Lessig, Social Meaning, supra note 182, at 1001.
edge the rationality of conformance to stylized rituals only for the sake of piety and grief. The Hindu widow’s piety is a matter of attitude, not of action for consequences; and her grief, like grief over death everywhere, is a wholly retrospective affair. All of this is antithetical to consequentialism’s prospectively oriented conception of practical rationality as the optimization of preference satisfaction.

Lessig might respond with a concession that he has represented the Hindu widow as one who optimizes the satisfaction of her preference for the expression of Hindu values. This is an admittedly imperfect reduction of her practical reasoning, the argument goes, but it is a reduction that will facilitate an enlightening economic analysis. But if this is Lessig’s argument, then we should note that it has two flaws.

First, expressive actions cannot be optimized in the way that Lessig seems to suppose. To say that the Hindu widow grieves in the Hindu way in order to optimize the satisfaction of her preference for the expression of Hindu values implies that, at some point of diminished marginal returns, the Hindu widow will cease to abide by Hindu beliefs and practices regarding death, and she will do so for this optimizing reason. This is false. The Hindu widow’s traditional expressions of grief are motivated by her grief and her piety, neither of which is a matter of the promotion of consequences, and neither of which is subject to the law of diminishing marginal returns. It is never the case that a person who engages in the expression of a value will cease to do so at some point of diminished marginal returns, because expressions of value are not motivated by their re-

262. The claim that piety is a matter of optimizing for the afterlife radically misunderstands religion. Granted that Hinduism conceives of the afterlife in terms of reincarnation, with a hierarchy of better and worse outcomes, it does not follow that Hindu piety is a matter of strategic choice. Compare Christianity, which features a promise of heavenly rewards, but denies that these rewards can be achieved by strategically good behavior. See Edwards, supra note 219, at 162-72. Is Hinduism really less like Christianity and more like game theory in this regard? 263. See Swanton, supra note 257, at 37-39 (distinguishing between the “consequential dangers of corruption or betrayal of desirable personal traits,” and “the rationality of a desire not to dishonor or be false to oneself or others” as a matter of personal identity and integrity); see also Foot, supra note 58, at 207 (arguing that virtuous action need not be described or evaluated in terms of better or worse states of affairs).
turns. To package the Hindu widow's beliefs for economic analysis in the way that Lessig does is to falsify them descriptively, and to dismiss them normatively. In the case of the Hindu widow, Lessig almost seems to recognize this fact, but he never quite comes to terms with it.

Second, if Lessig supposes that his analysis, like other economic analyses, has normative force, then he does not claim to have produced an imperfect, but serviceable, representation of the Hindu widow's practical reasoning. Instead, he claims to have reduced her practical reasoning to its sound essentials. That is, he claims that to eliminate retrospective and non-optimizing reasons for action from the model of practical rationality is to eliminate only reasons that are irrational or arational; that the prospectively oriented, optimizing reasons that remain constitute the core or essence of sound practical rationality; and that the economic analysis that rests on this construction of practical rationality therefore will generate accurate predictions.

264. Expressions of value are affective-cognitive responses to particular objects, and are meaningful within local systems of value. They serve, in part, the need to maintain a consistent identity, which is also served by means of one's deliberations on ends. On motivation as an affective-cognitive response, see supra text accompanying notes 246-47. On local systems of value, see infra text accompanying notes 271-79. On identity and deliberations on ends, see infra text accompanying notes 280-91.

265. Regarding the application of efficiency norms to the question of Hindu widowhood, Lessig writes:

But note an important and latent incoherence in the very notion of applying these norms to changes in social meaning. For when we begin to tinker with social norms or social meanings, we also begin to undermine the perspective from which one can really choose whether he or she is "better off" in the changed world over the unchanged world. For if a large part of who someone is is the sum of these constructions of social meanings—the set of practices or understandings that guide and constitute her—then this act of changing social meanings is an act of changing the individual herself. And if [this is] an act of changing an individual, what is the coherence in the claim that "the individual" is better off with the change than without it?"

Lessig, Social Meaning, supra note 182, at 1003.

266. He continues:

Resolving these questions is fundamental, but beyond the scope of this Article. In my view, they will have no simple resolution. For purposes of this Article, I will assume that the identity of individuals does not change as these constructions proceed, and hence, that it is coherent to speak of it being "efficient" to change certain meanings.

Id.
and sound policy prescriptions. But this claim to have reduced practical reasoning to its sound essentials surely requires a supporting argument. Why are retrospective and non-optimizing reasons less essential to sound practical reasoning than prospective, optimizing reasons? To reply that practical rationality just is a matter of the prospective optimization of value would beg the question.

These arguments against Lessig's treatment of Hindu widowhood can be made against any "social meaning" treatment of nonconsequential valuation or reasoning, including Kahan's analysis of deterrence. Nonconsequential rationality and action cannot be reduced to consequentialist terms by the simple expedient of supposing that the agent is engaged in the optimization of his preference for nonconsequential value. The optimization of nonconsequential value is an incoherent idea. If Kahan actually wishes to recognize expressive actions as rational actions, then he cannot maintain the economist's premise that all rational action is aimed at the optimization of preference satisfaction.

3. Intrinsic, Incommensurable, and Local Value

The third, fourth, and fifth features of nonconsequential value and practical reasoning that Kahan recognizes are closely related to one another.

Kahan recognizes intrinsic value. We value things such as people, animals, communities, and ideas for their own sake, and not merely as instrumental means to the attainment of favor-

267. To concede that this reduction has not been achieved would be to concede that economics cannot achieve the degree of normative force that economists have always attributed to their science; see Savage, supra note 221, at 20-21 (comparing the normative force of rational choice theory to the normative force of logic); see also Hollis & Sugden, supra note 224, at 6 (stating that Savage provided "what is still generally regarded as the most satisfactory statement of the theory of rational choice").

268. One common argument for the economist's construction of practical rationality is not an analytical or conceptual one, but a normative argument. That is, the economist maintains that his reduction of value and practical reasoning to the optimization of preference satisfaction is a sound reduction because it serves the ends of a liberal legal order better than nonconsequentialist alternatives. But this contention is unfounded. Nonconsequentialist ethics can support a liberal legal order, and consequentialism is normatively inferior to these alternatives because it tends to promote the commodification of all value. See Anderson, supra note 249, at 158-67.
able feelings in us.\textsuperscript{269} For example, I hope that my wife survives and flourishes long after my death has deprived me of any pleasure in her presence, because I value her being, quite apart from my responses to her. Kahan acknowledges that friendship involves this kind of value in one of his arguments, quoted above, for an expressive theory of alternative sanctions.\textsuperscript{270} If I offer a friend five thousand dollars instead of the evening that we had planned to spend together, and explain to her that the opportunity cost for my time that evening has risen to six thousand dollars, then I will have repudiated a vital part of our friendship. I will have said that I value her presence extrinsically, as a preference of mine that is to be coordinated with my other preferences for my optimal satisfaction; rather than intrinsically, for her sake or for the sake of our friendship itself.

Kahan recognizes local systems of value in the same example, when he acknowledges that my repudiation of the intrinsic value of my friend and our friendship occurs against a background of social norms.\textsuperscript{271} To speak of intrinsic value is not to say that value is a metaphysical property of things or that it exists in the abstract, apart from the agent who is engaged in valuation. Quite the contrary: valuation is a human practice. As Joseph Raz has argued, the intrinsic value of a person, idea, or animal lies "in its being a constitutive part of a valuable form of life,"\textsuperscript{272} where "form of life" means not only social practices and institutions, but also associated beliefs and attitudes.\textsuperscript{273} Elizabeth Anderson calls these local systems of value "decision frames."\textsuperscript{274} Lessig and Kahan recognize these local systems of value under the name of "social meaning."\textsuperscript{275}

The recognition of intrinsic value and local systems of value leads us to a fifth feature of nonconsequential value and practical reasoning, which also can be discerned in Kahan's illustra-

\textsuperscript{269} See Joseph Raz, The Morality of Freedom 177-78 (1986) (distinguishing intrinsic from instrumental value).
\textsuperscript{270} See Kahan, supra note 4, at 597.
\textsuperscript{271} See id.
\textsuperscript{272} See Raz, supra note 269, at 178.
\textsuperscript{273} See id. at 311.
\textsuperscript{274} See Anderson, supra note 249, at 23-26.
\textsuperscript{275} Lessig, Social Meaning, supra note 182, at 951-55.
tion of expressive rationality in friendship. Kahan recognizes the incommensurability of value when he acknowledges that "giving her the money and sharing my time with her are not interchangeable." This incommensurability is implicit in the notion of value's being intrinsic within local systems of value.

Suppose that I love both music and science, and that I admire the achievements of both Bach and Darwin. If you ask me to say whom I admire more, and I am unable to choose, you might think that my indecision reflects the fact that Bach and Darwin are so close in value that I am unable to distinguish their relative positions. This is not quite right, however. One can imagine a Darwin who achieved insights into genetic theory in addition to evolutionary theory, and this imagined Darwin would be more admirable than the real Darwin. Because the imagined Darwin is more admirable than the real Darwin, one should be able to say that the imagined Darwin is more admirable than Bach, even if Bach and the real Darwin are too close in value to distinguish. But it makes no more sense to say that the imagined Darwin is more admirable than Bach than it does to say that the real Darwin is more admirable than Bach. Bach and Darwin are incommensurable in value.

276. Kahan, supra note 4, at 597.
277. This example is offered by Elizabeth Anderson. See ANDERSON, supra note 249, at 55-56.
278. As this example shows, incommensurability has nothing to do with the difficulty of determining relative preferences, the difficulty of making hard choices, or the phenomenon of rough equality of options. See RAZ, supra note 269, at 328-35 (distinguishing and describing the relationship between equality of value and incommensurability of value). The failure to make this distinction can lead to the facile conclusion that minor elaborations of preference theory can address the philosophical problems that incommensurability raises. See, e.g., Richard A. Epstein, Are Values Incommensurable, or Is Utility the Ruler of the World?, 1995 UTAH L. REV. 685, 694-98 (arguing that apparent incommensurabilities can be resolved into interdependent utilities). The incommensurability of the options in the example is due to the fact that the valuations in question are intransitive: Darwin is not worse than Bach and the imagined Darwin is better than Darwin, but it does not follow that the imagined Darwin is better than Bach. See RAZ, supra note 269, at 324-26 (describing intransitivity as "the mark of incommensurability"). The phenomenon of incommensurability thus undermines the economist's assumption that revealed preferences are reliable indicators or representations of value. See id. at 325 ("In the study of people's actual valuations of different options, reliance on transitivity of this kind often leads the researcher to find either irrationality, or hidden preferences
Incommensurability such as this is a feature of values that are intrinsic within different local systems of value. Darwin and Bach are each valued within a system of value that is separate and distinct from that within which the other is valued: music in the case of Bach and science in the case of Darwin. The local values of music—variety, invention, the elegant development of a complex theme, emotional depth—are different from the local values of science: the breadth of the phenomena to be explained, the elegance of the hypothesis, the resistance of the theory to falsification, and so on.\textsuperscript{779}

To put the matter another way, the incommensurability of value, like intrinsic value, is an implication of the local systems of value that Lessig recognizes under the name of "social meaning." This creates an insoluble difficulty for Lessig and Kahan if they wish to stay within a law and economics framework. None of these three features of value— intrinsic value, incommensurable value, localized value—can be reconciled with economics. Fundamental economic concepts such as marginal utility and strategic choice under conditions of uncertainty depend upon the wholly incompatible premises that value is a function of personal preference, that value is commensurable across alternative goods, and that value can be optimized on a global scale. Kahan can recognize one of these two families of ideas about value, but not both.

\textsuperscript{779} The economist might argue that the notion of admiration is the problem in the example. He might say that admiration is a matter of opinion, not fact; and matters of opinion are arational. \textit{See} Epstein, \textit{supra} note 278, at 688-89 (equating incommensurability with the arational subjectivity of matters of pure taste). To assert that admiration is an arational response, however, is to ignore our ability to articulate the grounds for admiration in each sphere of activity. My admiration for Darwin is not a mere taste—it is not at all like my preference for chocolate over vanilla—because I can intelligibly debate Darwin's achievements with a fan of Galileo. We can argue about the achievements of each man in terms of the breadth of the phenomena that he explained, the elegance of his hypotheses, the resistance of his theories to falsification, and so on. In other words, admiration is an affective-cognitive response that constitutes a valuation with its own set of appropriate occasions and manifestations. \textit{See} GAUS, \textit{supra} note 241, at 25; \textit{supra} notes 245-57 and accompanying text.
4. Deliberations on Ends

Finally, a sixth feature of nonconsequential value and practical reasoning is implicit in the five that I have already described and that, with varying degrees of explicitness, Kahan acknowledges. I noted first that Kahan repudiates the Humean conception of value and motivation when he argues that some law enforcement measures will be effective because they appeal to one's self-conception, and to the logical implications of a given action for one's self-conception—that is, Kahan recognizes that value and motivation have cognitive dimensions. Implicit in this point is another one that is worth separate consideration: that practical reasoning is not reducible to instrumental reasoning because people deliberate on their ends.

Economic analysis will often refer to the agent's ends as her ends "whatever they may be," because an analysis of instrumental reasoning takes the agent's ends to be a given. In instrumental analysis by definition focuses exclusively on the agent's chosen means and their efficacy; never on the nature of the ends to which the means are directed, nor on the deliberations by which the agent might have determined her ends.

The problem with this omission, of course, is that actual people do deliberate on their ends; that is, on the standing motivations that constitute character. For example, suppose that my children are grown and my wife and I have reached a point of financial security that will permit us to retire and travel. However, I discover that my parents' health and financial situation suddenly have deteriorated. They can survive without my help, but not comfortably. No course of means-ends reasoning will tell me what I want to do in this situation; but neither will my ends

280. See, e.g., Eyal Zamir, The Efficiency of Paternalism, 84 Va. L. Rev. 229, 250 (1998) ("Decisions and actions that are based on consideration of irrelevant information, disregard of obvious risks, shortsightedness, faulty recollection, computational errors, or overconfidence, or that are taken without rational analysis of the available data, are unlikely to accomplish one's aims, whatever they may be, or to maximize aggregate social utility." (emphasis added)).

be determined by noncognitive desire. The kind of practical reasoning in which I will engage will merit the name of deliberation, and my deliberations will be about the system of ends or set of standing motivations that constitute the person I am and wish to be.

When Kahan repudiates the Humean conception of practical reasoning, according to which practical reasoning consists only of instrumental reasoning toward the satisfaction of one's noncognitive desires, he implicitly recognizes the possibility of these deliberations on ends. The notion that preferences are not fixed, but instead can be shaped, is a rudimentary version of this idea. 282

However, the notion of preference shaping cannot do justice to deliberations on ends unless it is separated from the premises of economics. To begin with, deliberations on ends are deliberations on one's character; that is, they involve reasoning about one's motivations and about the valuations that constitute one's motivations. The primitive noncognitivism that underwrites the economist's theory of value declares this kind of reasoning about value to be impossible. A neo-Humean theory of valuation and motivation, such as Gaus's affective-cognitive theory, can describe deliberations on ends. It can do this not only because this theory acknowledges the cognitive dimension of valuation and motivation, but also because it describes in detail the affective-cognitive responses that constitute the valuations that are the subject of practical deliberations. The economist, in contrast, cannot analyze deliberations on ends. Because the revealed preference theory of value is premised on completed choice, it recognizes only yes or no choices, pro or con attitudes, and black or white evaluations. As a consequence, it cannot describe the complexity of genuine practical deliberations. Given the primitive nature of a preference, the notion of preference-shaping cannot adequately capture deliberations on ends.

282. See Kahan, supra note 4, at 603-04 (describing the role of preference shaping in deterrence theory); see also Dau-Schmidt, supra note 3, at 24-32 (analyzing various features of criminal law as preference shaping mechanisms); Lessig, Social Meaning, supra note 182, at 1020-24 (describing the shaping of preferences in connection with AIDS education).
Furthermore, deliberations on ends frequently issue in something other than the optimization of value. Instead, they issue in the adoption of attitudes, such as piety; or in acts that are expressive rather than done for the sake of consequences, such as ritualistic mourning. For example, I might decide to assist my destitute father and mother in order to honor them. This act cannot be reduced to consequentalist terms by the simple expedient of supposing that I am optimizing my preference for honoring behavior. This would imply that, at some point of diminished returns from my honoring behavior, I will cease to honor my parents or to act under this imperative, and that I will do so because of these diminished returns. This is not true, and when the economist supposes that it is true, he overlooks whole classes of attitudes and actions that are the subjects of deliberations on ends. Nothing in the notion of preference-shaping cures this defect in the economist’s understanding of practical reasoning.

Finally, deliberations on ends are constituted largely by deliberations on incommensurable value. The consequentalist is fond of pointing out that such deliberations are impossible by definition: a choice between incommensurables can only be resolved by a flip of a coin. This specter of a comprehensive arbitrariness in our practical reasoning is taken to be an argument in favor of value’s commensurability. We do in fact make difficult choices, the argument goes, which shows that we can bring our options under a single metric, albeit sometimes with difficulty. However, as we will see, the consequentalist’s attempt to force an election

283. See, e.g., Donald H. Regan, Authority and Value: Reflections on Raz’s Morality of Freedom, 62 S. CAL. L. REV. 995, 1062 (1989) (making a coin-flip analogy). The concept of local value answers this objection in part, because it explains how we make rational decisions in the absence of a global scale of value. This answer is not quite sufficient. The consequentalist replies that the inevitability of conflicts between local scales of value and our apparent ability to resolve such conflicts must imply a global scale of value—that is, that value is commensurable. See Epstein, supra note 278, at 693 (“If these choices are incommensurate, then how can we prevent the ultimate fragmentation and disintegration of any and all comparisons between alternatives?”). This is a mistake because it overlooks the virtues’ role as normative structures that evolved to deal with problems of localized value specifically, including conflicts between incommensurable value. See infra text accompanying note 305.

284. This argument confuses hard choices, which are attributable to rough equality of value, with incommensurable choices, which involve intransitivity. See supra note 278.
between practical anarchy and a belief in commensurability is founded on a false dilemma. It reflects nothing more than a failure to acknowledge the existence of normative structures that have evolved for precisely this purpose: to cope with the incommensurability of value in the context of deliberations on ends. The economist’s continued insistence on the commensurability of value in his notion of preference shaping ensures that these norms will be overlooked, and that the notion of preference-shaping will not adequately describe deliberations on ends.

In sum, Lessig’s social meaning theory is untenable, but not because these six features of value and practical reasoning cannot be shaped into a coherent theory. Elizabeth Anderson has drawn these elements together into a plausible pluralistic theory of value. Lessig’s difficulty is that, unlike Anderson, he does not recognize that these features of value and practical reasoning imply value pluralism and a cognitive dimension in motivation; or that, as such, they contradict the monistic, non-cognitivist foundations of economic analysis.

Remarkably, Kahan equates Lessig’s social meaning theory with Elizabeth Anderson’s expressive theory of value, and has described his own work as an expressive theory of punishment. But Kahan is clearly mistaken when he equates the work of Lessig and Anderson. Lessig’s social meaning theory never seriously questions law and economics orthodoxy. Neoclassical economics is premised on particular assumptions about

285. See infra text accompanying note 305.
286. See ANDERSON, supra note 249 passim.
287. See id. at 79-86 (arguing against hybrid versions of consequentialism like Lessig’s).
288. For example, these two sentences—one of which cites Anderson, and one of which cites Lessig—begin and end a short paragraph: “As the work of Elizabeth Anderson reveals, selecting actions that express appropriate social meanings is part of being rational. . . . If we want to explain behavior, and identify effective legal strategies for changing it, we must take account of the expressive dimension of reality.” Kahan, Social Meaning, supra note 184, at 610-11 (citing ANDERSON, supra note 249; Lessig, Social Meaning, supra note 182). In another article, Kahan begins a paragraph with the Lessig-like proposition that “[a]ctions have meanings as well as consequences.” Kahan, supra note 4, at 597 (citing Lessig, Social Meaning, supra note 182). Then, he gives an account of expressive rationality that is taken from Anderson. See id. (citing ANDERSON, supra note 249).
289. See Kahan, supra note 4, at 596-97.
value, practical reasoning, and normativity that bring it within the category of consequentialist ethics. Anderson’s expressive theory of value, in contrast, is explicitly nonconsequentialist—indeed, it is probably more accurate to describe her arguments as vigorously anticonsequentialist. Any resemblance between the work of Lessig and Anderson is quite superficial, and Kahan’s running them together is nearly inexplicable.

C. Unanswered Questions About Fault

It should be apparent at this point why Kahan’s “new path” analysis of fault ultimately will not take us anywhere we need to go. Kahan offers analysis of mistake and fault that is framed in terms of a sophisticated law and economics version of deterrence theory. In this deterrence theory, Stigler’s notion of “illicit utility” is an important element. But both Kahan’s version of deterrence theory and Stigler’s “illicit utility” concept appeal to a background morality that is never specified or defended. Kahan appears to think that Lessig’s social meaning theory can be used to connect the economic analysis of the criminal law to the background morality of the criminal law, whatever it may be. This hope is unfounded and certain to be disappointed. Social meaning theory succeeds only in undermining or repudiating the premises of economics itself.

This puts Kahan’s analysis of fault at a dead end. The questions that Yeager and I have proposed in response to Kahan’s analysis of mistake remain unanswered. What is the nature of the morality that is the background of the criminal law? How are we to understand, jurisprudentially, the claims that this morality makes on the law? How is it that the law comes to govern the internalization of this morality’s norms?

290. See ANDERSON, supra note 249, at 163-67 (delineating the ethical limitations of “market ideologies”).
291. The best explanation for this confusion is the fact that Lessig’s account of “social meaning” is in fact an account of the same phenomenon that Anderson describes as expressive rationality. The difference is that Lessig simply has not absorbed the full implications of this insight. He has failed to see that it implies a plural, rather than a monistic, conception of value, and that it therefore undermines the monistic premises of economic analysis.
Only a consistent nonconsequentialist theory of punishment can answer these questions. I will sketch such a theory in the next Part.

IV. A VIRTUE ETHICS THEORY OF PUNISHMENT

Criminal negligence was included in the Model Penal Code section on “Kinds of Culpability” only over the strong objections of those drafters who adhered to a strictly intentionalist construction of culpability. The intentionalists lost this battle, in part because the Code drafters deferred to the probable wishes of actual legislators, and to a powerful, if difficult to articulate, rationale for fault:

[M]oral defect can properly be imputed to instances where the defendant acts out of insensitivity to the interests of other people, and not merely out of an intellectual failure to grasp them. In any event legislators act on these assumptions in a host of situations, and it would be dogmatic to assert that they are wholly wrong.

In the matter of criminal negligence, the Code drafters made the correct choice. In this Part, I will explain why the imagined legislators’ rationale is correct as well, not only for negligence, but also for other instances of nonintentional fault; and not only for nonintentional fault, but also for intentional fault. This “insensitivity to the interests of others” is a basic concern of the criminal law. Its opposite, a sensitivity to the interests of others, is a rough, but fair, description of virtue. To recognize the relevance of virtue to fault helps us to answer the three questions that Kahan’s “new path” deterrence theory leaves unanswered.

A. Value, Practical Reasoning, and the Virtues

Even if Kahan, unlike Lessig, were to recognize the value pluralism that I described in the previous Part, he would need to

293. Id. (citation omitted).
294. Thus, I plead guilty to being one of those academics who treats deterrence theory as “a foil to be battered to prove that some other abstract ‘theory’ supplies the ‘truth’ about criminal law.” Kahan, supra note 1, at 2496.
take one more step before he could solve the problem of fault. He would have to appreciate the kind of norms that these features of nonconsequential value and practical reasoning imply.

The virtues—such as loyalty, courage, generosity, and so on—capture the full range of practical rationality for purposes of normative governance.\(^9\) This breadth is evident as soon as one considers the way in which these norms address us. First, the individual virtues and vices are framed as character traits because they are directed \textit{not only} at actions for consequences, in the way that consequentialist norms are, \textit{but also} at one's expressive actions, and at one's attitudes and dispositions to action.\(^6\) Second, the acquisition and maintenance of these character traits are parts of the construction of a consistent character and stable identity over time, which is a project that involves retrospective as well as prospective reasoning. By framing its norms around attitudes, dispositions, and expressive acts and by incorporating retrospective as well as prospective considerations, virtue ethics attains a breadth and subtlety in human governance that consequentialism characteristically lacks.

For example, suppose that I buy season tickets for the home games of my local baseball team. Midway through the season it is apparent to me that my team will lose the vast majority of its remaining games, just as it has lost the vast majority of the games that it has played so far. The cost of my season tickets is unrecoverable regardless of whether I go to the remaining games. According to economic theory, a "sunk cost" of this kind should be irrelevant to my decision whether to go to the remain-

\(^{295}\) Elizabeth Anderson recognizes this connection between expressive rationality and Aristotelian virtue ethics, even though she does not develop it fully. She writes: "No adequate interpretation of a way of valuing something can reduce its motivational component to a desire or preference that some states of affairs occur. They must be brought about in the right ways, by the right agents, in the right context."\(^E\)\(^{249}\), at 30 (citing ARISTOTLE, supra note 25, at 43); see also id. at 37 (recognizing that expressive norms, as noninstrumental norms, resemble virtue); Kosman, supra note 219, at 108-09 (describing Aristotelian virtue as action from proper motives, construed as emotional states).

\(^{296}\) My italics are meant to emphasize that consequences are not left out of account in virtue ethics; courage is a virtue that wins empires. Due to this fact, the dividing line between consequentialism and virtue ethics is not entirely clear, and depends in large part on which version of virtue ethics one has in mind. See Oakley, supra note 26, passim.
The economist will insist that my going to the remaining games simply because I bought season tickets is irrational behavior.

Now, suppose that my grandfather squandered a generations-old fortune and my father died young from the combined effects of poverty's stress and his superhuman efforts to leave me a small legacy. These additional facts do not change the economist’s analysis. To the economist, these historical items are no more relevant to my decision than my sunk costs. Consequentialism, as the name implies, is inveterately forward-looking. But if this is my family’s history, do I really have no genuine reason to go to the baseball game, as the economist contends? Do I really behave irrationally or arationally if I treat my family’s history as a decisive consideration against skipping the season’s remaining games?

The answer is no. I can rationally use my season tickets to go to futile late-season baseball games. I will go in order to repudiate my grandfather’s way of life and to honor my father. And these expressions and attitudes are neither arational nor merely idiosyncratic to me: identifiable norms stand behind them. My repudiation of my grandfather’s way of life expresses condemnation of an identifiable vice, profligacy, and like one more brick in a good wall, my act reinforces my own efforts not to fall into the attitude or set of attitudes that constitute this vice. My emulation of my father’s conduct is likewise part of an ongoing effort to maintain a particular attitude or disposition that is identifiable as the virtue of frugality. It is also an act that is expres-


298. This project of mine cannot be reduced to consequentialist terms. See supra Part III.B.2, notes 253-68 and accompanying text. I do not hope to promote frugality itself; nor will I be frugal only to the extent that doing so efficiently advances my well-being generically conceived. I only hope to be frugal with the constancy that honor to my father demands. I might abandon this project at some point, for some reason. But the economist cannot simply stipulate that, at some point along an indifference curve, frugality will cease to promote my wealth or general utility efficiently, and that, I would abandon frugality at this point, for this reason. Such a reduction of my practical reasoning is Procrustean. Given the salience to me of my retrospective deliberations on my attitudes, I can demand that the economist justify their omission from an account of my practical reasoning with more than a conclusory insistence that, if I act on these deliberations, then I will have behaved
sive in accordance with the requirements of honor—to honor one’s father being a noteworthy attitude-governing norm.\textsuperscript{299} The consequentialist’s contention that my behavior is irrational is only a tacit concession that his conception of practical reasoning is too narrowly focused to make sense of it.

The virtues capture the full range of practical rationality for a third reason in addition to their addressing dispositions and retrospective reasoning. As the foregoing example suggests, the virtues and vices govern deliberations on ends. Deliberations on ends are deliberations on the standing motivations that constitute one’s character. The character traits that constitute the virtues and vices describe these standing motivations. When I attempt to live according to an individual virtue, I am attempting to acquire a particular end or set of ends, and to incorporate these ends into my standing motivations. This fact points to an essential feature of virtue ethics: the idea that the virtues and vices are not precepts that govern conduct on the occasion of action;\textsuperscript{300} that they are instead guides to the development of the fine sensibilities of the \textit{phronimos}—the person of exemplary judgment who is, for this reason, truly virtuous.\textsuperscript{301} Virtuous people, after all, are made, not born.\textsuperscript{302} Before one can act from virtuous motives and from a fixed disposition of virtue, one must have knowledge of the virtues and of the good for human beings.\textsuperscript{303} One acquires virtue by the practice of this knowledge; by habituation to the model of a virtuous life through deliberate conformity to the features of such a life: the virtues.\textsuperscript{304} Consequently.

\textsuperscript{299} See \textit{Exodus} 20:12.

\textsuperscript{300} Aristotle tells us that the dictates of virtue cannot be reduced to rules, see \textsc{Aristotle}, \textit{supra} note 25, at 4, 35, 141, 157-63, 248-49, that one who acts rightly only because he follows rules is not truly virtuous, see \textit{id.}, at 38-39, 170-71, and that virtue is more akin to perception than to reasoning; see \textit{id.}, at 170-73; see also David Wiggins, \textit{Deliberation and Practical Reason}, in \textit{Essays}, supra note 219, at 221, 235-37 (noting that a person’s choice in a given situation “does not reside in a set of maxims or precepts”).

\textsuperscript{301} Aristotle described this exemplary practical reasoning as \textit{phronesis}, and equated virtue itself with \textit{phronesis}. See \textsc{Aristotle}, \textit{supra} note 25, at 171 n.60.


\textsuperscript{304} See John McDowell, \textit{Deliberation and Moral Development in Aristotle’s Ethics},
sequentialism, in contrast, cannot adequately describe deliberations on ends, much less give them an intelligible normative structure.

The virtues capture the full range of practical reasoning for a fourth reason. In the course of deliberations on ends, incommensurable values are likely to come into conflict. The virtues serve as a means to reconcile these conflicts. This reconciliation does not necessarily come about before the agent takes action on the matter that presents the conflict. The person who faces a conflict of incommensurable values will act when she must, according to her best lights and her present sense of her best self. Her practical reasoning on the act will occur before and after she has acted, when she reflects on the implications of her act for her character. The reconciliation of conflicts between incommensurable values is accomplished, if at all, in the course of this long-term project. The virtues play their distinctive roles in these deliberations on ends.305


305. The consequentialist complains that this guidance is insufficiently prescriptive because it does not tell us what to do in advance, and may never tell us what ought to have been done. Consequentialism’s ideal is a decision procedure that, in Rosalind Hursthouse’s phrase, “provide[s] clear guidance about what ought not to be done which any reasonably clever adolescent could follow if she chose.” Rosalind Hursthouse, Virtue Theory and Abortion, in VIRTUE ETHICS, supra note 26, at 217, 224. Hursthouse goes on to explain that virtue ethics begins from the premise that “such a condition of adequacy is implausible.” Id.

Acting rightly is difficult, and does call for much moral wisdom, and the relevant condition of adequacy, which virtue theory meets, is that it should have built into it an explanation of a truth expressed by Aristotle, namely, that moral knowledge— unlike mathematical knowledge—cannot be acquired merely by attending lectures and is not characteristically to be found in people too young to have had much experience of life. There are youthful mathematical geniuses, but rarely, if ever, youthful moral geniuses, and this shows us something significant about the sort of knowledge that moral knowledge is. Virtue ethics builds this in straight off precisely by couching its rules in terms whose application may indeed call for the most delicate and sensitive judgment.

Id. (citation omitted). Put another way, virtue ethics is simply more frank than consequentialism. The rules or decision procedures that consequentialism provides inevitably will conceal some play in the joints: rules and procedures are always adjusted in their application to actual cases and particular circumstances. Virtue ethics acknowledges this play from the outset, as a fundamental feature of practical rea-
This tells us why the consequentialist is mistaken when he argues that if value really were incommensurable, then our deliberations on hard choices would make no sense; that our best option would be to flip a coin.\textsuperscript{306} He overlooks the virtues' role as normative structures that have evolved specifically to deal with problems of local value,\textsuperscript{307} including incommensurability. To substitute the virtues' coping with incommensurability for consequentialist preference optimization does not ensure that conflict between values does not occur. But it does mean that our ability to resolve such conflicts does not imply a global scale of value. Conflicts between values are not conflicts between contemplated sets of consequences, which are to be resolved before action is taken. Conflicts between values occur within the person, in the construction of a character; and conflicts such as these are resolved in one's ongoing and substantially retrospective deliberations on ends. In the context of deliberations such as these, the notion of one's promoting a global, impersonal, and acontextual value in one's choices and actions is incoherent.

\footnotesize

\textsuperscript{306} See Regan, \textit{supra} note 283, at 1062.\textsuperscript{307} Martha Nussbaum describes Aristotelian virtue in this way:

The question about virtue usually arises in areas in which human choice is both non-optional and somewhat problematic. (Thus [Aristotle] stresses, there is no virtue involving the regulation of listening to attractive sounds, or seeing pleasing sights.) Each family of virtue and vice or deficiency words attaches to some such sphere. And we can understand progress in ethics, like progress in scientific understanding, to be progress in finding the correct fuller specification of a virtue, isolated by its thin or nominal definition. This progress is aided by a perspicuous mapping of the sphere of the grounding experiences. When we understand more precisely what problems human beings encounter in their lives with one another, what circumstances they face in which choice of some sort is required, we will have a way of assessing competing responses to those problems, and we will begin to understand what it might be to act well in the face of them.

B. The Place of Virtue in the Criminal Law

I have suggested that virtue ethics can explain fault, and that it can provide a justification for punishment. The foregoing description of virtue ethics might seem instead to have foreclosed this possibility. Virtues and vices are framed as character traits. Criminal codes, on the other hand, consist of rules. The place of virtues and vices in the criminal law is obscure. Furthermore, the law's rules seem to operate in a fairly straightforward way: they prescribe certain actions in order to bring about certain states of affairs. Consequentialism seems likely, on this ground alone, to offer the best account of the normative foundations of the criminal law.

These observations are superficial and misleading. We do not see the virtues and vices as such in operation in the criminal law. But we do see the broader features of this nonconsequentialist ethics come into play.

Let me begin with an example of these features' coming to the surface in a difficult criminal law question. Some women who are accused of murder seem to have a strong case for excuse when the victim is a man who has severely abused his killer. When such abuse is extraordinarily degrading to the woman, and when it persists over many years, a persuasive case can be made that the less drastic routes of escape are closed to her, in any realistic sense. The depression, impaired cognition, and loss of perspective that accompany prolonged stress may make her choice to kill perfectly understandable, perhaps even inevitable, in psychological terms. However, whether her circumstances and resulting condition so constrain her as to render her ineligible for punishment is a different, and more difficult question.  

In order to resolve this more difficult question, we will engage in a highly fact-specific inquiry into the defendant's deliberations on, and choices about, her ends. An older woman who chose to remain in an abusive relationship seems less likely to succeed with the battered-woman defense than a younger woman who was trapped in the cycle of abuse and dependence before she could appreciate her situation. The older woman seems less

likely to succeed because it is more likely to appear that she could have been more perspicuous in charting the course of her life; that she could have avoided the abusive relationship, or left it before escape became impossible for her. This is a question of the quality of the woman’s practical reasoning, a question of virtue. This example shows us where questions of virtue come into play in the criminal law: the jury’s decision turns in part on its assessment of the quality of the defendant’s deliberations on her ends.

I hasten to add that the distinction that I have just drawn between older and younger battered women is itself highly controversial, perhaps even offensive to some readers because it assumes much about the control that a woman can or ought to exercise over her fate, and about the ends that a woman ought to have. I have chosen this provocative example deliberately, in order to highlight an additional dimension of the jury’s deliberations on virtue. The mere possibility that the jury could make such a distinction shows that the proper ends of a woman, and the manner and extent to which she can and ought to determine them, are subjects of the jury’s deliberations. The jury deliberates, not only about the quality of the defendant’s deliberations on her ends, but also about the content of the norms that govern a person’s deliberations on ends. In other words, the jury determines the requirements of virtue itself.

A virtue ethics theory of punishment takes this pattern to be typical of the criminal law. The legislation of conduct rules and the adjudication of individual cases comprise two continuous, integrated phases of a single normative enterprise. The latter, adjudicative phase involves deliberations and a decision about sound practical judgment, or virtue. The earlier phase, the positive legislation of a criminal code, provides a starting point for these deliberations on virtue.

From this description of the criminal trial, one can generate explanations of the justification of punishment, the nature of wrongdoing, and the nature of fault. In addition, one can see the point of adjudication; that is, a reason to reject the deterrence theorist’s assumption that adjudication is only an administrative matter that has no independent significance in the theory of
punishment.\textsuperscript{309} As noted above, fault is both inseparable from wrongdoing and an affirmative, justifying reason for punishment above and beyond the bare violation of a criminal prohibition.\textsuperscript{310} The jury's decision about virtue is the functional manifestation of this complex relationship.

1. Wrongdoing

Wrongdoing in the criminal law is defined by its primary norms, which are stated in the form of rules for the governance of conduct.\textsuperscript{311} A virtue ethics theory of punishment conceives of the conduct rules of the criminal law in a way that is fundamentally different from that of either of the two orthodox theories of punishment. A deterrence theory conceives of the criminal law's conduct rules as prescriptions that possess a background of justifying consequences that are expected to flow from the acts prescribed.\textsuperscript{312} A retributive theory of punishment conceives of rules as prescriptions which carry their own justification; for example, a formal justification based on their accord with the categorical imperative,\textsuperscript{313} or a divine justification based on their origin in the word of God.\textsuperscript{314}

According to a virtue ethics theory of punishment, in contrast, the conduct rules of the criminal law are premised on ethical generalizations. First, let me explain what I mean by an ethical generalization. The ordinary conduct of life involves choice and action; and some deliberation on ends lies behind the choices and actions that are taken in any coherent life by a person who has a stable identity. Our deliberations, choices, and actions fall into recognizable patterns simply because the biological and social features of human life are relatively stable. We can take

\textsuperscript{309} See supra text accompanying notes 80-81.
\textsuperscript{310} See supra notes 136-68 and accompanying text.
\textsuperscript{311} See HART, supra note 7, at 6-8.
\textsuperscript{312} See FREDERICK SCHAUER, PLAYING BY THE RULES: A PHILOSOPHICAL EXAMINATION OF RULE-BASED DECISION-MAKING IN LAW AND IN LIFE 17-27 (1991) (describing rules as having background justification).
\textsuperscript{313} See KANT, supra note 48, at 140-41 (describing the principle of punishment as a categorical imperative).
\textsuperscript{314} See ISRAEL DRAPKIN, CRIME AND PUNISHMENT IN THE ANCIENT WORLD 272-75 (1989) (describing Shari'a, the law of Islam, including the criminal law, as of divine origin).
note of these patterns and draw conclusions about which deliberations, choices, and actions are better and worse. In other words, we can and do generalize our ethical experiences. The individual virtues and vices are one kind of ethical generalization. Each virtue or vice is, in Martha Nussbaum’s words, a “perspicuous mapping of the sphere of the grounding experiences.”

Virtues and vices are, as I have noted, framed as character traits, but we can map our ethical experience in other ways, too. In particular, rules can be seen to resemble the virtues in this respect, and so can laws. We need not view either rules or laws as the consequentialist or the deontologist or the religionist views them. We can see laws instead in the way that the virtue ethicist sees the virtues and vices: as the product of a process of ethical generalization. So to say that the virtue ethics theory of punishment conceives of the conduct rules of a criminal code as ethical generalizations is to say something like the following. In any reasonably healthy political system, legislation is likely to take as its models those judgments concerning common, problematic situations that are widely regarded as sound judgments. The generalizations of these sound judgments are familiar to us, and almost uncontroversial: that property ought to be secure; that life and bodily integrity ought to be preserved; that individual autonomy ought to be respected; and so on. A criminal code is a relatively detailed set of positively enacted ethical generalizations of this kind. Their violation constitutes criminal wrongdoing.

2. Punishment’s Justification

What difference does it make to view the criminal law’s conduct rules in this way? In consequentialist and retributivist theories of punishment, the theoretical source of the law’s rules indicates the justifying ground of punishment. A virtue ethics theory of punishment is not different from its rivals in this respect. The criminal law’s conduct rules are generalizations from sound judgments concerning choice and action, and sound judgment concerning choice and action—virtue, in a word—is the justifying ground of punishment. Punishment is warranted in

cases in which a person's actions do harm, but not only because harm has been done.\textsuperscript{316} The harm done is a manifestation of the actor's lack of virtue; an indication that his practical reasoning is faulty. He has not merely chosen the wrong means to his ends; he has chosen the wrong ends. We condemn and punish him for this failure.\textsuperscript{317}

To conceive of the criminal law's rules and of punishment's justification in this way does not require us to alter radically our view of the criminal law's ordinary operations. Virtue ethics is sometimes thought to be antithetical to law, both because its norms usually are framed as character traits and because Aristotle stresses the development of context-sensitive judgment over governance by rules on the occasion of action.\textsuperscript{318} But it is occasional governance, not rules themselves, that virtue ethics rules out, and the operation of legal rules is not limited to occasional governance. We become habituated to legal rules, as Hart, for one, recognized:

Very often when a person accepts a rule as binding and as something he and others are not free to change, he may see what it requires in a given situation quite intuitively, and do that without first thinking of the rule and what it requires. When we move a piece in chess in accordance with the rules, or stop at a traffic light when it is red, our rule-complying behavior is often a direct response to the situation, unmediated by calculation in terms of rules.\textsuperscript{319}

Virtue ethics is in fact premised in part on this function of legal rules. Aristotle writes:

To obtain the right training for virtue from youth up is difficult, unless one has been brought up under the right laws. To live a life of self-control and tenacity is not pleasant for most people, especially for the young. Therefore, their upbringing and pursuits must be regulated by laws; for once

\begin{itemize}
  \item \textsuperscript{316} See Huigens, supra note 6, at 1429-37.
  \item \textsuperscript{317} See id. at 1440-44.
  \item \textsuperscript{318} See ARISTOTLE, supra note 25, at 170-73 (stating that virtue is more akin to perception than to reasoning); see also id. at 4-5, 35-36, 38-39, 141-42, 157-63, 170-72, 248-50 (stating that virtue cannot be reduced to rules).
  \item \textsuperscript{319} HART, supra note 173, at 140.
\end{itemize}
they have become familiar, they will no longer be painful. But it is perhaps not enough that they receive the right upbringing and attention only in their youth. Since they must carry on these pursuits and cultivate them by habit when they have grown up, we probably need laws for this too, and for the whole of life in general.\textsuperscript{320}

To the extent that one finds Hart's description of legal obligation to be plausible, one ought to be open to the neglected theoretical possibilities of a virtue ethics legal theory, including a virtue ethics theory of punishment.

As this conjunction of Hart and Aristotle is meant to suggest, the notion of the internalization of legal rules is the point at which a sophisticated deterrence theory of punishment borders on a virtue ethics theory of punishment. The difference between the two theories lies in the role and significance that each of them assigns to the internalization of rules. In deterrence theory, Hobbes,\textsuperscript{321} Bentham,\textsuperscript{322} and Holmes\textsuperscript{323} continue to state the central case: the criminal as an instrumentally rational optimizer of preferences. In Hart's hands, the internalization of legal rules appears as a fortunate feature of human psychology that extends the reach, and thereby the plausibility, of deterrence theory.\textsuperscript{324} In contrast, virtue ethics places the internalization of legal rules at the center of its theory of punishment: it interprets the internalization of legal rules as habituation to virtue; and takes the habituation to virtue to be punishment's justifying reason for being.

\textsuperscript{320} ARISTOTLE, supra note 25, at 295-96; see also id. at 34 ("Lawgivers make the citizens good by inculcating (good) habits in them, and this is the aim of every lawgiver. . . .").

\textsuperscript{321} See HOBBES, supra note 77, at 44-45, 214 (describing punishment's purpose as the disposition of the will, and the disposition of the will as the product of instrumental reasoning about competing passions).

\textsuperscript{322} See BENTHAM, supra note 68, at 155, 174 (describing punishment as operating on the motive of self-preservation, but not in cases of unintentional action).

\textsuperscript{323} See supra note 27 and accompanying text.

\textsuperscript{324} See supra notes 170-78 and accompanying text.
3. Fault

The virtue ethics theory of punishment does not require us to alter radically our conception of the criminal law's ordinary operations, but the theory does offer at least one significant advantage over its dominant rival. The foregoing view of adjudication, of legislation, and of punishment's justification provides us with an explanation of fault that is superior to consequentialist constructions of fault.

I have argued that fault has a primary normative dimension; that it is an aspect of wrongdoing that pertains to the particular manner and circumstances of the wrongful conduct in actual cases. As a first step in the explanation of fault, let me give a functional account of this relationship between fault and wrongdoing. In the adjudication of every case, the primary norm that is alleged to have been violated is applied as a secondary norm by the jury. For example, the ex ante prohibition on theft becomes, in the hands of the jury, a standard by which to determine ex post whether the acts of the defendant are punishable as theft. This application of the primary norm as a secondary norm in adjudication involves more than a nominal change in the norm's function. When the jury applies the primary norm as a secondary norm in the course of its deliberations, the jury also re-casts the primary norm, explicitly or not, from its generalized rule-form into the concrete terms of the individual case. If we mean anything of substance by the phrase, "applies the law to the facts," then this is what we mean. When the jury interprets its instructions in light of the particular facts of the case; when it determines the facts in light of its instructions; and when it gives legal effect in a verdict to the facts that it has found, the jury necessarily frames the primary norm as a secondary norm in the concrete terms of the case before it.

This specification of the primary norm in terms of the particulars of the individual defendant's circumstances is a necessary part of the determination of fault. The conduct rule that

325. See supra notes 136-68 and accompanying text.
326. This point is a basic feature of the analysis of primary and secondary norms. See Dan-Cohen, supra note 30, at 628-29 (regarding the application of conduct rules as decision rules).
defines the offense is premised on an ethical generalization. The conduct rule thus presents an implicit demand that the accused should engage in sound practical reasoning in the circumstances in which such a crime might be committed. The requirement of fault, a secondary norm, entails a retrospective, adjudicative inquiry into this question; that is, into the virtue of the accused, where virtue refers not to rote compliance with a code of duties, but instead to the refined and perspicuous practical judgment of the phronimos. The question before the jury is whether the acts of the accused in the particular circumstances of the alleged crime displayed inadequate or flawed practical reasoning, including the deliberations on ends that have gone toward establishing and maintaining his standing dispositions.\(^3\) This determination by the jury turns on a comparison between the defendant's judgment in the relevant circumstances and a sound judgment in the relevant circumstances—the latter as evidenced by the applicable conduct rules, which are generalizations of sound judgments in the relevant sphere of human life and conduct.

The fault of the accused, then, is an inference that the accused lacks virtue; that his practical judgment is inadequate or flawed. This inference is drawn by the jury when it applies the primary norm as a secondary norm; that is, when it specifies the primary norm in terms of the desires, choices, and actions of a virtuous person in the circumstances of the accused, and then compares this standard to the actual conduct of the accused. This interpretation accounts for the primary-normative aspect of fault.\(^3\) Fault is conceptually and functionally inseparable from the violation of the primary norm that constitutes wrongdoing, because the jury specifies the primary norm when it draws the inference of fault.

Both fault and the adjudicative specification of the primary norm by which fault is determined are required by the fact that the inculcation of virtue is the criminal law's justifying purpose. Whether the accused's choices and actions are virtuous choices

---

327. In some cases, of course, this issue will be uncontroversial. For example, in the O.J. Simpson murder trial, it was clear that an unjustifiable homicide had occurred. The question for the jury was instead one of identity.
328. See supra notes 133-52 and accompanying text.
and actions depends upon the particular circumstances of the accused. As a result, no case can be decided under the criminal law until the ethical generalization of the conduct rule is returned by the jury, in the course of its deliberations, to the level of particularity from which the conduct rule originated, and at which virtue or its absence is manifest. This particular requirement of justice entails the inquiry into fault: the jury's comparison between the defendant's judgment in the relevant circumstances and the sound judgment that is implicit in the applicable conduct rule.

This adjudicative inquiry into fault brings the justification for the conduct rules' prohibition to bear upon the individual case of punishment in a justifying way. If the justifying purpose of punishment is the inculcation of sound practical judgment or virtue, and if the primary norm that one is accused of violating is a generalization of sound practical judgments in the field in which the crime occurs, then the fault of the accused lies in his practical judgment. 329 We have an affirmative, justifying reason to blame and to punish the accused—one that is distinct from the wrongdoing itself—in that his actions evince inadequate practical reasoning on the occasion of the act, which inadequacy in turn evinces a long-term failure of the accused to deliberate well on his ends and to assemble a set of ends that is congruent with the common good. 330 This inference is fault.

This connection between the justification of a prohibition and fault is the reason that the midcentury deterrence theorist's distinction 331 between the justification of punishment as a practice—by which he means the system of primary prohibitions—and the justification of instances of the practice—by which he means the adjudication of individual cases under secondary rules for the distribution of punishment—is inapplicable to the criminal law. Adjudication is not an insignificant administrative task relative to legislation. Legislation and adjudication are continuous phases of a single normative process that in-

329. This is the nature of the connection between character and fault that has frequently been noted. See Arenella, supra note 6, at 61; Kahan, supra note 2, at 144 (citing R.B. Brandt, A Motivational Theory of Excuses in the Criminal Law, in CRIMINAL JUSTICE, supra note 10, at 165, 176-77).
330. See Huigens, supra note 6, at 1458-62.
331. See supra text accompanying notes 72-81.
volves the complementary generalization and specification of ethical judgments.

Both intentional and nonintentional fault are determined in the way that I have described. In many, perhaps most, cases, the particularities of the accused, his circumstances, and his behavior include an intentional state regarding harm or a risk of harm, and this intentional state denotes fault. However, this intentional state does not constitute fault. Fault consists of a broader set of the facts that describe the way in which the accused has come to do wrong and the particular way in which he has done wrong. If this failure of practical reasoning is not indicated by an intentional state on the occasion of action, then we have an instance of nonintentional fault, but we do have an instance of fault.

C. The Rule of Law Objection

The emphasis that the virtue ethics theory of punishment places on the specification of the primary norm by the jury inevitably raises concern for the rule of law. The jury's specification of the primary norm might differ considerably from the legislature's authoritative statement of the primary norm, which is the only norm of which the wrongdoer had notice. The virtue ethics theory of punishment seems to be premised on ex post facto lawmaking.

Four points may be made in response to this objection. First, the jury's fixing the terms of the primary norm ex post is actually a relatively common feature in the criminal law. It has been deemed to be consistent with the rule of law when punishment is justified on consequentialist grounds. Some argument

332. Every crime that involves a reasonableness determination in connection with one of its elements involves an ex post facto determination of a primary norm. Criminal cases involving reasonableness determinations such as these include not only crimes that are premised explicitly on criminal negligence, but also any crime that, in its interaction with some other doctrine of criminal law—such as accomplice liability—results in what the Model Penal Code calls the equivalent of liability for negligence. See, e.g., MODEL PENAL CODE § 2.06 cmt. 6(b) n.42 (1985); see also supra note 114 (noting the Code drafters' concern over negligence liability if accomplice liability is premised on foreseeability).

beyond merely noting the ex post facto feature of a virtue-ethics theory of punishment is needed in order to explain why this feature is an insurmountable difficulty when the justification for punishment is offered in nonconsequentialist terms.

Second, we must distinguish between descriptive and normative claims. If it is the case that the justification of punishment implies an adjudicative determination of fault, and that this determination of fault entails the specification of the primary norm in an ex post facto fashion, then our abhorrence of ex post facto lawmaking does not tell us that the analysis is descriptively wrong. It tells us, instead, that we face a normative dilemma. In order to resolve it, we might wish to abandon the practice of punishment altogether. Or, we might wish to go about the practice of punishment in a way that mitigates the ex post facto problem at the cost of some loss of precision or detail in our determinations of fault. For example, we might wish to adopt an intentionalist construction of fault such as that stated in the Model Penal Code's section 2.02. But none of this would be to deny that it is the case that the justification of punishment implies an adjudicative determination of fault, and that this determination of fault entails specification of the primary norm in an ex post facto fashion.

Third, the jury is necessarily the final arbiter of the primary norm in every case, and the danger that this necessity presents is kept within rule-of-law bounds by ordinary doctrines of appellate review for sufficiency of the evidence. I have argued that fault is an aspect of wrongdoing that is determined by the jury

---

334. Sufficiency of the evidence review is always implicitly a review of the jury's interpretive construction of the governing norm. See, e.g., Commonwealth v. Pahel, 689 A.2d 963 (Pa. Super. Ct. 1997); see also Darryl K. Brown, *Jury Nullification Within the Rule of Law*, 81 MINN. L. REV. 1149, 1169-70 (1997) (arguing that some jury nullification consists of inevitable interpretive construction of law). In *Pahel*, the defendant was convicted of having "knowingly endanger[ed] the welfare of the child by violating a duty of care, protection, or support," 18 PA. CONS. STAT. ANN. § 4304 (1983), on evidence that she had failed to seek prompt medical attention for a child's broken nose. The appellate court left undisturbed evidence to the effect that "the amount of trauma needed to inflict a nasal fracture could have caused injury to the brain." *Pahel*, 689 A.2d at 964. Nevertheless, the appellate court reversed for insufficient evidence of risk to the child. *See id.* at 967. In other words, it rejected the jury's interpretation of the statute, under which the endangerment of a child encompassed a failure to treat a broken nose.
in the course of its application of conduct rules to particular facts. A delicate balance between governance by rules and the use of juries to attend to the particularities of individual cases is inherent in this theory. We have no reason to suppose that a virtue ethics theory cannot take advantage of the law as we know it in order to maintain this balance at a familiar point.

Finally, in defense of a virtue ethics theory of punishment, it is certainly worth recalling Aristotle's treatment of the rule of law. Aristotle insists on the rule of law, not of men, on the ground that only governance by reason can be impartial and even-handed. Laws must be clearly and definitely stated. Equity is to be resorted to only in case of law's failure, and the necessity of resorting to equity in specific cases does not cast doubt on the justice of law. Individual rights are to be protected under a sound constitution.

Aristotle identified justice and the lawful with virtue, but, as the foregoing features of the Aristotelian political ideal indicate, this is not to suggest an unfamiliar or sinister set of legal institutions. Instead, the ordinary institutions of law consist of right action that honors the virtue of justice. To adopt and to live according to rules is right, not because rules have good consequences or because a legal order is dictated by the categorical imperative, but because to create and to live under a legal order honors the virtue of justice. Therefore, to give the jury's fine-

335. See ARISTOTLE, supra note 25; at 128-30; THE POLITICS OF ARISTOTLE 145-47 (Ernest Barker trans., Oxford Univ. Press 1946) [hereinafter ARISTOTLE, POLITICS]. 336. See ARISTOTLE, supra note 25, at 240-41. 337. See id. at 141-42. 338. See FRED D. MILLER, JR., NATURE, JUSTICE, AND RIGHTS IN ARISTOTLE'S POLITICS 131-39 (1995). 339. See ARISTOTLE, supra note 25, at 112-14; ARISTOTLE, POLITICS, supra note 335, at 145-50. 340. See MILLER, supra note 338, at 133-34 (citing, inter alia, ARISTOTLE, POLITICS, supra note 335, at 3-5, 118-20, 331; ARISTOTLE, supra note 25, at 128-30, 293-96). For example, were I to find myself in a position to rule over a group of others, the virtuous thing to do would be to decline to rule by decree and instead to initiate constitutional deliberations that involve the entire community and lead to the creation of a legal order. To rule by decree would be to deny the others the opportunity to participate in political life, and to develop their practical wisdom by doing so—ultimately to my detriment as well as theirs. Similarly, we act virtuously in the creation and maintenance of our actual legal institutions. 341. To institute a legal order is the right action, ultimately, because of the kind of
grained determination about virtue precedence over the rules in every case—that is, never to hold that the jury has exceeded its authority in the specification of the primary norm—is not a necessary or desirable part of a virtue ethics theory of punishment.

CONCLUSION

The foregoing virtue ethics theory of punishment answers the three questions that Kahan's analysis of mistake and fault leaves open.

What is the nature of the morality that underlies the criminal law? It is not a morality at all, if by "morality" one means a discrete set of duties that parallels our legal duties. The foundation of the criminal law is instead an ethics of virtue, which claims no transcendent authority because it consists only of the generalization—in rule or character-trait form—of those judgments about common, problematic experience that, over time, we take to be sound judgments that are worthy of emulation.

How are we to understand, jurisprudentially, the claims that this morality makes on the law? The process of ethical generalization occurs in the ordinary course of legislation in a representative democracy, and results in laws that embody the determinations of sound practical judgment, or virtue. In the trial of a criminal case, the determinations of virtue that are implicit in the law's conduct rules are unpacked and applied to the case of the accused at the level of particularity at which the rules originated.

How is it that the law governs the internalization of community norms? Hart noted that we come to follow legal rules unconsciously. Virtue ethics simply provides a different gloss on this familiar phenomenon. If community norms begin in particular instances of sound practical reasoning, then both the law's enactment of these norms in legal rules and its concrete and particular re-enactment of these norms in adjudication are devices that assist us—along with the family, school, and religion—in the acquisition of sound judgment in practical matters of com-

See ARISTOTLE, supra note 25, at 15-16, 214-16, 263-64.
mon concern. In sum, the justifying purpose of the criminal law is the inculcation of virtue.

To say that the justifying purpose of the criminal law is the inculcation of virtue is not to say that deterrence has no role in the law. Just as instrumental reasoning cannot be omitted from one's conception of practical reasoning, so the deterrence of crime by an instrumental appeal to the potential criminal—"Do this, and you will go to jail"—cannot be omitted from one's conception of the criminal law. But this simple, instrumental transaction must be knocked down from the conceptual eminence that it has occupied for deterrence theorists from Hobbes to Bentham to Holmes. Modern deterrence theorists, beginning with Hart and continuing with Kahan, have increasingly come to recognize that valuation and motivation are not the simple matters that these theorists' intellectual forebears supposed them to be, and that no purely instrumentalist and consequentialist account of the criminal law can possibly do it justice. The question is how to integrate these insights into the theory of punishment.

This Article has argued that it is not enough to acknowledge only nonconsequential values and a cognitive dimension in motivation. If the object is to integrate these phenomena into legal theory, then one must pay attention to the kind of normative system that these phenomena imply. One can then integrate this normative system into legal theory.

To take this approach in the theory of punishment enables one to answer the central riddle of the theory of punishment—the nature of fault—and to resolve the quandaries that have grown up around this riddle, including: the inculpating effect of a mistake of fact; the persistence of criminal liability for negligence; the inculpating effect of voluntary intoxication; and the question of which battered women can justly be convicted of murder.

The principal obstacle to this approach to the theory of punishment seems to be the name of the normative system that nonconsequential value and practical reasoning imply: virtue ethics. Misconceptions about virtue abound, ranging from the sophisticated, such as the notion that rules are antithetical to virtue, to the silly, such as the notion that virtue consists of a rigid adherence to a code of duties. In fact, virtue ethics is a
powerful, subtle, and ancient tradition in philosophical ethics. It is the true new path in the theory of punishment.