On Commonplace Punishment Theory

Kyron Huigens
Kyron.Huigens@chicagounbound.edu

Follow this and additional works at: http://chicagounbound.uchicago.edu/uclf

Recommended Citation
Available at: http://chicagounbound.uchicago.edu/uclf/vol2005/iss1/12

This Article is brought to you for free and open access by Chicago Unbound. It has been accepted for inclusion in University of Chicago Legal Forum by an authorized administrator of Chicago Unbound. For more information, please contact unbound@law.uchicago.edu.
On Commonplace Punishment Theory

Kyron Huigens

Some criminal law scholars have argued that the undertheorization of the criminal law is a good thing.\(^1\) As I have argued elsewhere, this seems unlikely, because the dangers of theoretical procrusteanism are well outweighed by the potential gains from a comprehensive theory of punishment.\(^2\) The topic of this paper, however, is not the undertheorization of the criminal law, but its incorrect theorization. At the risk of giving offense, I will criticize the commonplace punishment theory that shows up in ordinary briefs, judicial opinions, and—far too frequently—in the law reviews. By “commonplace punishment theory,” I mean a more or less well defined set of truisms that serve on an ad hoc basis whenever a bit of theory seems to be called for in ordinary criminal law practice, criminal justice policymaking, or scholarly articles on criminal law doctrine.

To illustrate the idea of commonplace punishment theory and the kind of danger it harbors, let me give an example that is not among the three major examples that I discuss below. Between the late 1970s and mid-1980s, criminal sentencing in the United States underwent a revolution, with the federal government and many states moving from indeterminate sentencing to determinate sentencing systems.\(^3\) The predominant feature of

---

\(^1\) Professor of Law, Benjamin N. Cardozo School of Law.


these new schemes was the severe curtailment of judicial discretion in punishment.⁴

The idea of disparity should have received a good deal more critical attention than it did from the judges and legislators who proposed and enacted the new schemes. Disparity in and of itself is not objectionable in sentencing. A murderer and a shoplifter should receive disparate sentences because they have committed crimes of disparate severity. What the reformers actually meant to address was inequality in sentencing. Murderers should receive sentences that approximately match the sentences that other murderers receive; likewise for shoplifters compared to other shoplifters; and so on for all other offenders. But inequality is not what the reformers were after, either. If we sentence all shoplifters to be drawn and quartered, then these offenders have received equal sentences. But clearly something has gone wrong. The proper objection is that the shoplifters have not received proportionate sentences. It was disproportionality—not disparity and not inequality—that the reformers ought to have focused on.

This is especially clear when one considers the asymmetrical relationship of equality and proportionality. Proportionate sentences are by definition equal; if each offender has received the punishment that he deserves, then all similarly situated offenders have been treated similarly. The converse is not true. As our pair of drawn and quartered shoplifters demonstrates, not all equal punishments are proportionate. At best, the reformers of twenty-five years ago were using equality as a proxy for proportionate punishments, without quite realizing how imperfect a proxy it is.

This paper will examine three important and prominent examples of commonplace punishment theory. According to innumerable briefs, judicial opinions, and academic articles, punishment requires proof of culpability, especially if one adopts a retributive theory of punishment.⁵ Furthermore, according to

⁴ Schulhofer, 128 U Pa L Rev at 737 (cited in note 3) (“More commonly, reform proposals treat broad judicial discretion as the principal evil.”).

⁵ See, for example, Tison v Arizona, 481 US 137, 149 (1987) (“The heart of the retribution rationale is that a criminal sentence must be directly related to the personal culpability of the criminal offender.”); Enmund v Florida, 458 US 782, 801 (1981) (stating that the retributive end is ensuring that “the criminal gets his just deserts”); Randy Hertz and Robert Weisberg, In Mitigation of the Penalty of Death: Lockett v. Ohio and the Capital Defendant's Right to Consideration of Mitigating Circumstances, 69 Cal L Rev 317, 369 (1981) (“The retributive judgment is a determination of the degree of punishment and suffering that is appropriate for or proportionate to the moral culpability of the offender and his offense” (citing Jack P. Gibbs, The Death Penalty, Retribution, and Penal Policy, 69 J Crim L & Criminol 291, 293–96)).
commonplace punishment theory, culpability is paradigmatically a matter of intentional states of mind. At least three things are wrong with this picture. First, retribution as a function of punishment is insufficient to define a theory of punishment because the retributive function can be theorized in a number of ways. Second, culpability is an ambiguous term that should be replaced by the more accurate terms “fault in wrongdoing” and “fair candidacy for punishment.” And, finally, intentions are not paradigmatic of criminal fault. On the contrary, the seeming outlier among fault concepts, negligence, is closer to the paradigm of criminal fault.

I. THE CONFLATION OF FUNCTION AND THEORY

One of three pervasive confusions in commonplace punishment theory is the tendency to talk in terms of “the retributive theory of punishment,” or “the deterrence theory of punishment,” or, less frequently, “the rehabilitative theory of punishment.” This way of talking conflates two different things: the ends of punishment and theories of punishment. Because of this confusion between ends and theories, lawyers, legislatures, and courts are often unable to answer specific questions about punishment that they need to answer, such as how sentencing ought to be done.

We ought to distinguish three separate parts of an adequate theory of punishment. First, there are many conceivable ends, or aims, of punishment: retribution, general deterrence, specific deterrence, incapacitation, rehabilitation, public catharsis, and the internalization of social norms, to name the most prominent. Second, there are many issues that a theory of punishment should be able to explain, beginning with the question of punishment’s justification, but extending well beyond it: the definition of wrongdoing, the analysis of fault, the ground of the excuses, the nature of justification defenses, the meaning of proportionality in sentencing, and so on. Third, given that many of these questions about just punishment are moral, not legal,

---

6 As Justice Robert Jackson articulated it:

The contention that an injury can amount to a crime only when inflicted by intention is not a provincial or transient notion. It is as universal and persistent in mature systems of law as belief in freedom of the human will and a consequent ability and duty of the normal individual to choose between good and evil.

questions, there are at least three major traditions in philosophical ethics on which we might want to draw for such explanations: utilitarianism or, more broadly, consequentialism; the deontological morality most prominently advocated by Immanuel Kant; and the virtue ethics of Aristotle and his philosophical heirs. Looking at punishment theory from this perspective, three important points can be made. First, punishment's justification by a particular end is not what defines theories of punishment. Much of commonplace punishment theory turns on the notion that a theory of punishment must take some one end of punishment as the primary justification of punishment, and that which end is chosen for this purpose defines each theory. For example, the so-called retributive theory of punishment says that punishment is justified by retribution, the so-called deterrence theory of punishment, says that punishment is justified by deterrence, and so on. But a more careful mapping shows that the field is more complex. For example, we can make connections between retribution as an end, consequentialism as a theory, and justification as an issue in need of explanation: consequentialism will deny that retribution has any justifying force—except perhaps indirectly through the consequence of public catharsis, which then becomes the consequentialist's reformulation of the end of retribution. Connections of this kind are concealed by conventional wisdom, which offers a choice between retribution as our theory and justification of punishment or deterrence as our theory and justification of punishment. At a minimum, to map things more carefully makes the point that whether or not retribution or deterrence is taken to justify punishment depends on the moral theory that is brought to bear on that function.

Second, this mapping reminds us that there is more work to be done by a theory of punishment than to answer the question about punishment's justification. The consequentialist needs to explain not only punishment's justification, but also the distinction between the practice of punishment and an instance of pun-

---

7 Consider Marcia W. Baron, Philip Pettit, and Michael Slote, *Three Methods of Ethics: A Debate* (Blackwell 1997) (discussing the competing merits of consequentialism, deontological morality, and virtue ethics).


9 See, for example, Russell L. Christopher, *Deterring Retributivism: The Injustice Of “Just” Punishment*, 96 Nw U L Rev 843, 874 (2002) ("In other words, though deterrence theories may be unjust in justifying the intentional punishment of innocents, retributivism is unjust in exposing the innocent general public (innocent future crime victims) to a greater risk of victimization.").
ishment, particularly as the distinction bears on the nature of criminal wrongdoing and the meaning of proportionality in sentencing. Otherwise, it might appear that the "deterrence theory of punishment" authorizes imposing draconian penalties on scapegoats for the sake of maximum deterrence.¹⁰

Third, to distinguish end from theory can clarify each of the things so distinguished. The deontological theory of punishment justifies punishment by appeal to retribution: to take retribution on a wrongdoer is an unconditional duty.¹¹ This duty, of course, has nothing to do with vengeance. Retribution as an end of punishment is the restoration of a proper balance—of benefits and burdens¹² and of dignity¹³—between the offender, the victim, and society. Given that Kant's deontological moral theory is not the prevailing moral theory in our society, the deontological theory of punishment has never been very influential with lawyers, lawmakers, or judges. It hardly follows from this, however, that retribution ought to be rejected as an end of punishment or that retribution cannot be adduced as a justifying reason to punish under some other moral theory. And yet, retribution as an end of punishment has been chronically misconstrued as vengeance, and also has been rejected, in baby-bathwater fashion, along with "the retributive theory of punishment."

Does this confusion matter? Of course it does. The most pressing issue in contemporary constitutional and criminal law is the validity of determinate sentencing schemes under the recent Supreme Court decisions Apprendi v New Jersey,¹⁴ Blakely v Washington,¹⁵ and United States v Booker.¹⁶ This very lively controversy over the limits imposed on sentencing by the Sixth Amendment jury trial guarantee has exhibited both of the confu-

¹⁰ The answer to the scapegoating objection to the consequentialist theory of punishment's justification involves a distinction between justifying punishment in individual cases and justifying legal systems of punishment within which only a legal justification for punishment need be offered in individual cases. Under the latter kind of argument for punishment's justification, the legal system would bar scapegoating as a rational side-constraint on punishment within that legal system. For versions of this argument, see, for example, H.L.A. Hart, Prolegomenon to the Principles of Punishment in Punishment and Responsibility: Essays in the Philosophy of Law 1, 5–11 (Oxford 1968); H.L.A. Hart, Legal Responsibility and Excuses, in Punishment and Responsibility at 111; John Rawls, Two Concepts of Rules, 64 Phil Rev 3, 3–4 (1955).


¹⁵ 124 S Ct 2531 (2004).

¹⁶ 125 S Ct 738 (2005).
sions I mentioned just above. Some of the determinate sentencing systems, especially the Guidelines, impose draconian sentences. Any number of commentators have concluded that, because the sentences they impose are draconian, these schemes represent the triumph of a retributive theory of punishment. But while the draconian sentences handed out under these schemes might give vent to the public's urge to take revenge on criminals, they do not serve the aim of retribution. The imposition of a draconian penalty can upset the proper balance between society, the victim, and the offender just as surely as the crime itself has done. The point of reference for retribution is desert, and at least part of a draconian punishment is undeserved. Draconian punishments might often serve the end of revenge, but they will rarely, if ever, serve the end of retribution. Furthermore, the imposition of draconian sentences is not a victory for any theory of punishment, even if we grant the conflation of vengeance with retribution. To recognize the pursuit of retribution or vengeance as an end of punishment is insufficient to constitute a "retributive theory of punishment," or even a "vengeance theory of punishment." To identify one end of punishment hardly justifies punishment in the absence of some moral theory that can tell us whether and why it does so; and does not even make a start on answering other outstanding questions such as the nature of wrongdoing and its relationship to the excuses.

Second, consider a prominent disavowal of retribution as a "theory of punishment" that was mistakenly thought to entail the disavowal of retribution as an end of punishment. Stephen Breyer understands punishment well enough to have been the principal author of the Federal Sentencing Guidelines, and also well enough to recognize that retribution has to do with desert, not with vengeance. But this is what he said about the theory of punishment that informs the Guidelines:

[S]ome students of the criminal justice system strenuously urged the Commission to follow what they call a

---


18 For example, in United States v Galloway, 976 F2d 414, 416-17 (8th Cir 1992) (en banc), the prescribed sentence for the defendant's theft was 21-27 months. Under the Federal Sentencing Guidelines' "real offense" sentencing scheme, Galloway was sentenced to serve 63-78 months based on seven other uncharged and unproven thefts.

"just deserts" approach to punishment. The "just deserts" approach would require that the Commission list criminal behaviors in rank order of severity and then apply similarly ranked punishments proportionately. For example, if theft is considered a more serious or harmful crime than pollution, then the thief should be punished more severely than the polluter.

... 

[T]he Commission soon realized that only a crude ranking of behavior in terms of just deserts, based on objective and practical criteria, could be developed. Although guidelines motivated by a just deserts rationale would be cloaked in language and form that evoke rationality, using terms such as "rank order of seriousness," the rankings would not, in substantive terms, be wholly objective.

... 

Faced . . . with those who advocated "just deserts" but could not produce a convincing, objective way to rank criminal behavior in detail, . . . the Commission reached an important compromise. It decided to base the Guidelines primarily upon typical, or average, actual past practice. 20

To resort to "average, actual past practice" to determine specific punishments is to disavow not only the "just deserts" approach to punishment, but also just deserts as an aim of punishment. Sentences based on "average, actual past practice" might or might not be deserved sentences. Indeed, a common complaint made against the Guidelines in action is that the sentences they impose are not deserved: they are frequently far too great in proportion to the wrongdoing actually done. No desert-based check was imposed on the Guidelines' proposed sentences, due largely to the skepticism about desert-based sentencing that Breyer expresses regarding the rank ordering of offenses. 21 This skepticism itself was unwarranted, but that is a topic I will address separately below.

The Guidelines drafters might have rejected a retributive theory of punishment and still pursued the aim of imposing de-

20 Id at 15-17.
served sentences. Had they realized this, perhaps they would have been a little more clear about the differences between disparity, equality, and proportionality in sentencing that I described in the Introduction. The aim of proportionality and the aim of just deserts are obviously quite closely related.

II. THE AMBIGUITY OF "CULPABILITY"

Consider the following statements: "Culpability consists of purpose, knowledge, recklessness, or negligence," and "The truly insane defendant should not be punished because he is not culpable." What if, as in most cases of insanity, the defendant acts purposely? Say D suffers from the delusion that he is Jack Ruby, that V is Lee Harvey Oswald, and that as a matter of historical necessity, D must shoot V to death. Is D "culpable" or not? You might be tempted to waffle: D's purpose was not a "real" purpose or a "culpable" purpose, because it was insanely arrived at. But how can it be seriously maintained that D did not have the objective, crazy or not, to cause the death of an entity that D knew to be a human being? This is not, after all, the same case as that in which X suffers from the delusion that Y is the angel Beelzebub, X is the Angel Gabriel, and that it is X's angelic duty to destroy Y—the corporeal form that Beelzebub has assumed as a disguise. In that case, X's insanity does deprive him of the objective to cause the death of an entity that X knows to be a human being. So why equivocate on "purpose" or "knowledge" in the Oswald/Ruby case? Why not just admit that one of the central terms in commonplace punishment theory is ambiguous?

One kind of "culpability" turns on the definition of a criminal offense because it is an aspect of criminal wrongdoing. For example, if a person shoots and kills a figure in the woods, believing that it is a deer, and the figure turns out to be a human being instead, then the killing is not a murder. A human being is dead, but the shooter's belief that he was shooting at an animal precludes the requisite finding that he had a purpose to kill an entity that he knew to be a human being. If the victim was walking in the woods during hunting season dressed in brown clothing, then the shooter probably would not be reckless or negligent either, precluding liability for manslaughter. The kind of culpability that is missing in these cases—which include the Beelzebub/Gabriel case—is commonly called mens rea, but a better term for it is "criminal fault."
A different kind of culpability, or its absence, can be discerned in cases in which the elements of an offense, including fault, are present. The Oswald/Ruby example falls in this category, but of course others can be imagined. Suppose, for example, that a first grader is angry with a classmate. He brings the family handgun to school the next day and shoots the other child—but, fortunately, not fatally. When the first grader is asked why he shot his classmate, the child answers honestly that he wanted to hurt the other child. All the elements of assault are present here, but no criminal liability will be imposed, because a six year old child is not subject to the criminal law at all. The feature missing from these cases usually is referred to as culpability, or blameworthiness, but both of these terms are also commonly applied to cases in which fault is missing. The precise term for this kind of defense is the “absence of responsibility,” but responsibility is also confusingly used to refer to a judgment of liability. In order to avoid these ambiguities, it is better to refer to the kind of culpability that is missing from cases such as Oswald/Ruby as “fair candidacy for punishment.”

There is a simple test that can determine whether a particular fact is being invoked as a matter of fault or fair candidacy. Notice that we refer to the absence of culpability in cases in which a person is not a fair candidate for punishment, such as an assault by a child, and yet it seems odd to refer to the mere fact that one is eligible for punishment as culpability. In contrast, in cases in which fault is at issue, we talk not only about the absence of culpability, but also—on the positive side, so to speak—about the varying degrees of culpability. The test, then, is: take the exonerating or mitigating version of the fact that bears on culpability and try to think of the inculpating or aggravating converse. If there is a plausible converse, then you are dealing with fault. If there is no plausible inculpating or aggravating converse, then you are dealing with fair candidacy.

For example, my not seeing that the figure in the woods is a man, not a deer, serves to exculpate me or to mitigate my offense, whereas any level of awareness that the figure is a man will inculpate me and, furthermore, will aggravate the offense by degrees corresponding to the degrees of awareness. In contrast, age usually works as a fair candidacy consideration. If a young

22 See Kyron Huigens, Rethinking the Penalty Phase, 32 Ariz St L J 1195, 1238–39 (2000) (considering various terms for the issue of fair candidacy for punishment).
23 Id.
24 Id.
person has been convicted of a serious crime, it makes sense to use age as a mitigating factor. But there is no corresponding aggravator. A judge might deny the mitigator of age, saying “You were old enough to know better,” but this is only to deny the mitigation of youth. She will not go on to say: “You’re lucky you’re only 20. If you were 30, I’d sentence you to life, and if you were 40, I’d sentence you to death.” Because age is being used here as a fair candidacy consideration, increasing the age does not aggravate. Likewise, being middle aged or elderly is not an aggravating factor in any death penalty statute, even though youth is a constitutionally required mitigator.

Does the confusion inherent in this bit of commonplace punishment theory matter in the real world? Of course it does. The Supreme Court has tied itself in knots over the question of the proof of mitigating factors in death sentencing. Must the jury be unanimous in finding mitigating factors, or will it suffice if one juror finds a mitigating factor? Does each mitigating factor have to be found unanimously, or will it suffice that the jury is unanimous on there being some mitigator, even if they disagree on which one is present? The Court has said that the jury need not be unanimous in finding a mitigating factor. Otherwise, the death penalty will be imposed if only one juror refuses to find

25 It might be objected that increasing age does aggravate in cases of child sexual abuse. Many statutory rape provisions grade the rape according to the discrepancy in age between the perpetrator and the victim. Charles A. Phipps, Children, Adults, Sex and the Criminal Law: In Search of Reason, 22 Seton Hall Legis J 1, 62 (1997). But this just shows that facts are not inherently either fault or fair candidacy considerations. It would be odd if they were. On the contrary, what is at issue here is the use, not the nature, of certain facts such as age. The question is whether the fact is being invoked for one purpose or another, and indeed, in cases of statutory rape, the age of the defendant is invoked as a fault consideration instead of a fair candidacy consideration. The discrepancy in the age of the perpetrator versus the victim is an aspect of the wrongdoing that makes it a worse or less bad instance of the offense, or that determines the threshold issue of whether to punish or not. For example, sex between a minor and a person over 18 sometimes is not a crime at all if the discrepancy in their ages is less than 4 years, as in the case of an eighteen year old boy and a fifteen year old girl. See Rev Code Wash 9A.44.079 (2004) (criminalizing sexual intercourse with a person “who is at least fourteen years old but less than sixteen years old and not married to the perpetrator and the perpetrator is at least forty-eight months older than the victim”); compare Idaho Code 18-6101(1) (2005) (defining rape as any sex with a female less than eighteen years old). A related point: as my example indicates, fault and fair candidacy are both relevant to just sentencing, though only fault is relevant to the definition of offenses. But this should not be cause for confusion. Fault and fair candidacy are relevant to sentencing in different ways: fair candidacy is a consideration that pulls on proportionality in only one direction, downward, whereas fault pulls both upward and downward.


27 McKoy, 494 US at 435.
any mitigator. Furthermore, if all jurors must agree on a single mitigator, then twelve jurors can agree that there is mitigation, but the jury will find no mitigation if one juror points to a different mitigator from the other eleven. The Court's ruling matches our intuitions of justice, but it implies, falsely, that each single juror, and not the jury as a whole, is the "sentencer" in capital cases. More important, it has the defect of departing from a general requirement of unanimity in criminal cases—a requirement that forces deeper deliberations by juries, and that compensates (in terms of respect for, and the authority of, the decision) for jury secrecy and for the jury's undemocratic and unrepresentative composition.

The root of the difficulty is the assumption that the defendant bears the burden of persuasion on mitigators. This distribution sounds plausible, and the Court has never questioned it. But notice that to put the burden of disproving mitigators on the government would solve the unanimity problem, because the risk of non-unanimity would fall on the government. In a case in which the defendant has presented a prima facie case on a mitigator, the jury will find mitigation if 11 jurors believe there is mitigation, or even if one juror believes there is mitigation. Where the defendant has presented a prima facie case on several mitigators, the jury will return a verdict of mitigation even if they are unable to agree on which of the mitigators is present.

To distinguish fault in wrongdoing from eligibility for punishment provides an argument for placing the burden of persuasion on mitigators on the government. All mitigators bear on culpability, but it is worth noticing that many of these issues bear on fault in wrongdoing instead of eligibility for punishment. (This point is clearer when one recognizes, as is argued in Part III, that fault in wrongdoing is not confined to intentional states

---

28 See Mills, 486 US at 473–74 (describing practical effects of a policy whereby jurors must be unanimous in finding mitigation).
29 Id.
30 McKoy, 494 US at 466 (Scalia dissenting).
32 Of course, many mitigators in death sentencing do pertain to the offender's capacities and to the question of fair candidacy for a particular punishment such as death. See Huigens, 32 Ariz St L J at 1204 (cited in note 22). A separate argument is needed to distribute the burden of persuasion on these mitigators to the government. See id at 1275–76 (arguing that the burden of persuasion on fair candidacy mitigators should follow the distribution of fault-based mitigators instead of the contrary distribution for full defenses premised on fair candidacy considerations).
of mind.) The convicted murderer attempts to show that she was a minor participant in the crime, or that the victim played a role in his own death. These questions bear on the degree of the wrongdoing, and not on the capacities of the defendant or on any other question relevant to fair candidacy. That being the case, the burden of disproving these mitigators ought to be on the government, just as it is in complete defenses that pertain to the presence or absence of fault, or to the proof of wrongdoing generally. A defendant's prima facie case of mistake of fact negates fault in the offense, and the government bears the burden of persuasion on mistake because it must prove all the elements of the offense, especially fault. A defendant's prima facie case of self defense negates the wrongdoing implied by the prosecution's proof of the offense, and the government bears the burden of persuasion on self-defense for that reason. So it is with the defendant's prima facie case that he was a minor participant or that the victim played a role in the homicide. In each case, the issue is fault, and the government must always carry the burden of persuasion on issues of fault. Had the Court recognized this feature of mitigators years ago, the confusion in its death sentencing jurisprudence could have been substantially reduced.

III. THE EQUATION OF FAULT AND INTENTIONAL STATES OF MIND

"Mens rea" is said to be an essential component of just punishment. Strict criminal liability is typically defined as punishment imposed without proof of mens rea, and it is thought by some to be unjust for precisely this reason. And yet, a literal "mental thing" is not required for punishment. Implied malice is still a viable doctrine in some places, and criminal liability for extreme negligence is recognized almost everywhere. The prevailing view is that implied malice and negligence are substitutes for true mens rea, and that cases premised on implied malice or negligence show that the "mens rea requirement" need not be taken literally. Some have argued, however, that liability premised on such "objective" criteria is always unjust. 

33 See, for example, Morissette, 342 US at 250–63 (finding that there are deep historical traditions of injury only amounting to crime when inflicted by intention).

34 People v Combs, 101 P3d 1007, 1031 (2004) (defining implied malice murder as not requiring proof of an intent to kill and allowing defendant to be sentenced to second degree murder when he lacked such intent).


36 Id at 239. See also Hart, Punishment and Responsibility at 136 (cited in note 10).

gence is said to be nothing but strict liability by another name, and implied malice is portrayed as a relic of a barbaric past.

Some minor confusion occurs here because criminal law theorists use the terms "subjective" and "objective" in ways diametrically opposed to their colloquial meanings. Colloquially, "subjective" means a matter of opinion, an inarticulable judgment that borders on the arbitrary and that cannot in a literal sense be right or wrong. "Objective" in a colloquial sense is the opposite of this: a matter of demonstrable fact about which one can be right or wrong. Punishment theorists use "subjective" to refer instead to the state or condition of an actor, so that to have a purpose or to have some knowledge is said to be a subjective state of mind, and liability premised on such mental states is said to constitute "subjective" criminal liability. "Objective" criminal liability rests on considerations external to the criminal actor—on his failure to meet a standard of due care or the moral depravity of his conduct. Now for the tricky part—punishment theorists tend to argue that "subjective" criminal liability (in the technical sense) is preferable because it is more "objective" (in the colloquial sense). An actor's state of knowledge or his aims can be determined as a matter of fact. Conversely, "objective" criminal liability (in the technical sense) is said to be objectionable because it is too "subjective" (in the colloquial sense). Due care is not a matter of fact, and the depravity of someone's conduct seems to be purely a matter of opinion.

This terminological confusion is a minor annoyance. It makes much of the scholarly literature inaccessible to the lay person because the inversion of terms is naturally confusing. But criminal practitioners and judges tend to crack this inadvertent code without much difficulty. The larger, more troubling confusion arises from writers taking seriously the colloquial meanings of "subjective" and "objective," and, worse, the arguments about just and unjust punishment that use these terms. Many writers contend, in essence, that objective criminal liability is unjust because it is too subjective. Objective—or, in the term preferred here, non-intentional—criminal fault is said to be essentially ar-

40 See, for example, Ronald Gainer, The Culpability Provisions of the Model Penal Code, 19 Rutgers L J 575, 591 (1988) (decrying the "mixture of moral precepts and emotional interplays that underlaid our previous assessment of mens rea").
bitrary because, unlike states of mind such as purpose and knowledge, matters such as negligence and malice involve value judgments, are prescriptive instead of descriptive, or are normative instead of factual.

In short, commonplace punishment theory holds that criminal liability ought to be limited to prohibited acts accompanied by discrete intentional states of mind because only such mental states can legitimately constitute criminal fault. The Model Penal Code’s inclusion of criminal liability for negligence was opposed on this ground.\textsuperscript{41} The modern law of criminal fault exhibits the dominance of a few intentional states elements over a variety of non-intentional criteria, and this is usually portrayed as a triumph of modernity and clear thinking over the vagueness and injustice of a benighted past. From this prevailing point of view, intentional states are paradigmatic of criminal culpability, negligence is an anomaly, and doctrines such as implied malice should be hunted by reformers into well-deserved extinction. Furthermore, much of this latter work remains to be done. A number of criminal law doctrines can be understood as premising criminal liability on non-intentional fault: felony murder, depraved heart murder, transferred intent, accomplice liability, unreasonable mistake, and the ban on the use of voluntary intoxication as a defense to crimes of recklessness.

There are a number of difficulties here. First, the foregoing account of criminal fault has been the prevailing view among criminal law theorists for a half-century or more, and it seems to be the view of the most prominent theorist who actively advises on the drafting of criminal codes.\textsuperscript{42} Yet most of the non-intentional fault doctrines survive, year after year, decade after decade. At some point, the persistence of these doctrines counts as a reason to change our theory instead of our practice. If a doctrine such as felony murder is widely held to be unjust, and yet it proves impossible to eradicate, then its persistence may well be due to something other than prejudice, perversity, or rationality defects in legislatures. It may be that the doctrine has been protected by a core of logic and legitimacy that immunizes it against attempts to reform it out of existence. A more promising task


\textsuperscript{42} I am referring to Paul Robinson. See, for example, Paul H. Robinson, Michael T. Cahill, and Usman Mohammad, \textit{The Five Worst (And Five Best) American Criminal Codes}, 95 Nw U L Rev 1, 17, 43 (2000) (critically commenting on the use of negligence and malice in defining criminal offenses).
than abolition, then, would be to uncover that legitimate core of felony murder. Having done this, one could then try to properly account for—and then eliminate—the doctrine's unjust and inessential aspects. The problem with felony murder, it turns out, is not that it is an "objective" or non-intentional fault doctrine, but that it is an inflexible—and therefore vastly over- and under-inclusive—structure of rules for the determination of non-intentional fault. The problem with felony murder, ironically, is too heavy an emphasis on legality and not enough sensitivity to the context and circumstances that constitute criminal fault.

A second difficulty with the prevailing view of criminal fault goes back to Blackstone. The theory behind the intentional states account of fault is that if a crime is committed unintentionally, then the defendant's will has not been engaged and his act is not one for which he can be held responsible. But this confuses volition (the capacity for choice that is a condition of responsibility) and intentionality (the awareness of something, such as the nature of one's conduct or the results it will produce). To act unintentionally is not to act without volition. The workman who negligently tosses a hundred pound bale of shingles from a roof onto a pedestrian below chooses to toss the bale of shingles even if he is unaware, though he should be, of some of the consequences of doing so. Voluntariness in gross is a requirement for just punishment, but this is different from the exercise of choice with respect to each and every aspect and consequence of one's action, and different again from the intentionality that might accompany those multiple choices. It is true that responsibility and therefore criminal liability require a conscious moral agent—a genuine actor—who is free to choose and capable of choosing. It does not follow that criminal liability requires proof of either a

---

43 Blackstone wrote:

All the several pleas and excuses, which protect the committer of a forbidden act from the punishment which is otherwise annexed thereto, may be reduced to this single consideration, the want or defect of will. An involuntary act, as it has no claim to merit, so neither can it induce any guilt: the concurrence of the will, when it has its choice to do or to avoid the fact in question, being the only thing that renders human actions either praiseworthy or culpable.


44 *Martin v State*, 17 S2d 426, 426 (Ala App 1944) (holding that an intoxicated person carried out of a tavern and deposited in the road cannot be convicted of appearing drunk in a public place).
choice or an intention regarding each and every aspect of the offense.  

The third difficulty with the prevailing intentional states view of fault is that cramming the fault involved in the offense into the intentional-states box has notorious adverse consequences. If we try to define the worst murders by reference to an intentional state called premeditation, then we will have to treat the elderly husband who kills his terminally ill and gravely suffering wife as a worse murderer than the one who tortures his victim to death while remaining so intent on torture that he overlooks the risk of death. If we try to define rape by reference to intentional states that must rise at least to the level of advertent recklessness, then we will have to acquit the man who is too callous, immature, self-involved, and stupid to perceive that a woman's struggle means refusal, not passion.  

The final difficulty with the prevailing view of fault is the metaethics that prompted it: specifically, the notion that objective or non-intentional fault doctrines are somehow lacking in realism or rationality. The underlying idea is expressed in the colloquial sense of "subjective." Judgments of morality and ethics are held by many to be "subjective" in this sense, as incapable of being either true or false. Legal doctrines that descend from, appeal to, or run parallel to moral judgments are deemed suspect for this reason, because they seem unmoored from reason and the natural world. Obviously, a legal doctrine based only on "subjective" value judgments in this way would threaten injustice. But the metaethics that undergird this view—the conception of action, value, normativity, and practical rationality—are seriously outmoded.  

The metaethics behind the hostility to non-intentional fault trace most immediately to the metaethics of A.J. Ayer circa 1945 and his arguments in Language, Truth and Logic.  

There, Ayer argued that value judgments are nothing more than emotional expostulations. To say that murder is bad is to say only "Boo murder!" and to say that charity is good is to say only "Yay charity!" Such expostulations are neither true nor false. They are merely reports of subjective states that are literally meaningless.

---

45 Another, even weaker, argument for requiring intentionality as a condition of just punishment is that if the defendant acts unintentionally he has not chosen to break the law. But choosing to break the law cannot be a requirement for just punishment unless we wish to recognize a universal defense of mistake of law.  


47 Id at 107–09.
(here using “subjective” in its technical sense, but in the context of the argument that, historically, is most likely responsible for the colloquial sense of “subjective” as arbitrary and arational).

Ayer’s metaethics, however, have no currency among philosophers today, and have not had any since the 1960s. As Peter Geach noted in a 1965 article, we use evaluative terms and normative propositions in several ways, each of which is meaningful and logical on its own terms.\(^4\) We might say, as a descriptive proposition: “If murder is bad, then it is bad to murder my brother.” Ayer would equate this with: “If boo murder, then boo the murder of my brother.” But the latter statement is an assertion, not a description, and while it might be a mistake to infer the assertion from the description or vice-versa, the assertion and the description are both meaningful, intelligible statements.\(^5\) Whether these intelligible, logical statements have any real-world content is another, more difficult, question,\(^6\) but it is clear at least that these statements are meaningful. This view of value judgments should go a long way toward relieving our anxiety about non-intentional fault. The fear that non-intentional fault doctrines are unjust comes from a fear that they are arbitrary, and the fear that they are arbitrary comes from a fear that they descend from, appeal to, or run parallel to moral judgments that are meaningless. But this is not true of moral judgments, and it is not true of non-intentional fault doctrines that track or derive from moral judgments.

An important effect of these clarifications is to reverse the picture of intentional states as paradigmatic and negligence as anomalous. In negligence and the other non-intentional fault doctrines, fault is found not in a discrete mental state, but in a broader set of facts surrounding the offense. These fault doctrines are paradigmatic because criminal fault generally is but an aspect of wrongdoing. Criminal fault is an inference, drawn in the course of the adjudication of wrongdoing, to the effect that the practical reasoning of the defendant is deficient. This assessment of the quality of the defendant’s practical reasoning is not limited to the defendant’s reasoning in connection with the offense—even if “in connection with” is given a very broad con-

---


\(^6\) It has been argued that our evaluative and normative statements are intelligible and meaningful, but necessarily always false. J.L. Mackie, *Ethics: Inventing Right and Wrong* 30–35 (1977) (stating this “error theory”). But see Hurley, *Natural Reasons* at 278–90 (cited in note 49) (refuting six different formulations of Mackie’s error theory).
struction. It extends, in addition, to an assessment of the defendant's set of standing motivations, or ends—to their acquisition, development, maintenance, and ultimate issuance in the alleged offense. From an opposite perspective, criminal fault is an aspect of criminal wrongdoing. That is, the manner, circumstances, and specifics of the individual instance of wrongdoing alleged against the defendant are the subject matter of the adjudicative assessment of the quality of his practical reasoning that I have just described.\(^5\)

The imposition of responsibility for the state of one's ends is part of a system of just punishment in part because the proper disposition of individual actors' ends is one of the objectives of punishment.\(^5\) This conception of fault is an aretaic, or virtue-based, conception of fault in part because it supposes that one's ends are the subject of rational deliberation, and that one can, therefore, be held responsible for the state of one's ends.\(^3\)

This is a novel and complex way of looking at fault, and interested, perplexed, or offended readers are invited to explore the footnote references. But a more concrete account also might help. Consider a criminal case in which identity is not at issue and which is sufficiently contested to have withstood the enormous exogenous pressures toward a guilty plea that are built into our present system. In a well-tried case of this kind, the lawyers on both sides will convey a highly specific, fact-rich account of the events leading up to and surrounding the alleged crime. In the course of deciding the case, the jury will determine whether the actions of the defendant violate the nominal terms of the criminal prohibition.

But the conscientious jury will not leave matters there, and the fact that we encourage the construction of dense, detailed accounts of the crime at trial indicates that we do not want the jury to leave matters there. When punishment is at issue, we want the jury to do fine-grained justice in light of the particulars of the individual case. Part of this fine-grained determination is an inquiry into the quality of the defendant's practical reasoning. If the prohibition has been violated, then the practical reasoning of the defendant in the circumstances of the offense apparently conflicts with the sound practical reasoning implicit in the prohibition.

---


The difficulty is that the prohibition is a generalization, whereas the defendant is a particular individual whose conduct occurred in particular circumstances. It may be that his nominal violation of the terms of the prohibition does not, on closer examination, run contrary to the practical judgments that went into making up the prohibition. It may be that a case of nominal wrongdoing does not actually evince the failures in practical reasoning and deliberations on ends that constitute, in part, the wrongdoing at issue. If this is so, then the nominal violation of the prohibition is not a true or genuine instance of wrongdoing, and punishment would not be justified. That is, the defendant could not be justly punished because he was not at fault.

This broad, theoretical picture of criminal fault obviously does not describe most actual adjudication of fault. One countervailing consideration accounts for most of the variance. A free form inquiry into the quality of the defendant's practical reasoning as it bears on whether he deserves punishment would threaten core rule-of-law values. The consistency and predictability of such inquiries obviously could not be guaranteed. To the extent that the fault inquiry entailed an express specification of the criminal prohibition, it would entail retrospective and retroactive lawmaking by an undemocratic and non-representative body deliberating in secret. Accordingly, we write rules and standards for the adjudication of fault. Furthermore, we have found that carefully defined intentional states of mind serve this purpose well. They capture much of what we mean by fault—it is worse to break a person's arm purposely than to break it recklessly or inadvertently—and intentional states of mind are (at least superficially) descriptions of states of affairs that juries can treat as matters of fact, and not as raw normative judgments. Non-intentional fault doctrines often seem unjust—think of felony murder—not because they embody untenable kinds of fault or do not reference fault at all, but because they are poorly framed rules and standards for the determination of fault. In general, rules about fault framed in terms of intentional states of mind formalize and regulate the jury's deliberations on fault in most cases.

This analysis implies, however, that criminal fault is not identical to any intentional state of mind. Rather, intentions are indicators of criminal fault that we incorporate into the positive law definitions of criminal offenses for rule-of-law reasons that are exogenous to fault (though not to just punishment). Nor are intentions paradigmatic instances of fault. Criminal fault is a conclusion about the quality of the defendant's practical reason-
ing that is drawn, within rule-of-law bounds, from the complete set of facts adduced at trial. From this perspective, non-intentional fault doctrines such as negligence and implied malice are paradigmatic of fault.

Does the confusion of commonplace punishment theory on this point have adverse real-world consequences? Of course it does. I have already mentioned one of the most troubling. Following the Model Penal Code, most American criminal codes have a provision stating that criminal liability for serious offenses requires proof of an intentional state of mind of at least advertent recklessness, and a corresponding provision restricting the use of criminal negligence to minor offenses.\(^5^4\) Often, however, either the state legislature or the state courts will require a mistake regarding non-consent in rape to be a reasonable mistake in order to acquit. This is contradictory.\(^5^5\) A genuine mistake of fact logically precludes finding fault premised on an intentional state of mind, whether the mistake is reasonable or not. To require a reasonable mistake regarding non-consent in order to acquit is to permit conviction for rape based on unreasonableness—negligence—regarding non-consent.

Furthermore, to resolve the contradiction in a way that accords with the supposed necessity of intentional-states fault produces implausible results. As noted above, if the alleged rapist is callous, immature, self-absorbed, and stupid, and if he has taken the victim’s struggle as a sign of passion instead of as a refusal for that reason, then requiring proof of an intentional state of mind regarding non-consent will acquit him. Under a Model Penal Code system, a mistake of fact serves to negate intentional states fault. If it never occurs to an alleged rapist that the victim is non-consenting, then he necessarily does not act intentionally—neither purposely, knowingly, nor recklessly—with regard to the “non-consent” element of rape. He will be acquitted precisely because he is callous, immature, self-absorbed, and stupid. Obviously, the contradiction should be resolved in the other direction—in the direction of accepting negligence as a legitimate basis for conviction of serious crimes and for severe punishments.

This analysis has significant practical implications. The challenge in the drafting of criminal offenses is not to cram the fault involved in all offenses into the intentional-states box. The

---

\(^5^4\) See Joshua Dressler, Understanding Criminal Law 332 (Matthew Bender 3d ed 2001).

\(^5^5\) See Paul H. Robinson, Criminal Law Defenses § 62(c)(2) at 253, 253–54 n 26 (describing such an inconsistency in Texas and New Jersey law).
real challenge is to formulate non-intentional fault doctrines that do not have the rule-of-law defects that are so prominent a feature of doctrines such as felony murder, transferred intent, strict liability, and so on. The challenge is to find the right balance between, on one hand, the rule-of-law values served so well by the intentional-states construction of criminal fault, and, on the other hand, an important competing value. This competing value is fine-grainedness, defined as a low level of under- and over-inclusiveness in our rules of criminal liability, relative to a background justification of moral desert. Alternatively, we could define it as a high degree of congruence between our legal judgments of desert and our moral judgments of desert. If legality is the rule of law as a law of rules, then the pursuit of fine-grainedness requires a relaxation of legality. The pursuit of legality, on the other hand, entails a loss in fine-grainedness and at least the risk of a loss in the credibility and prestige of the criminal justice system. The requirement of reasonable mistake regarding non-consent in rape is a perfect example of courts and legislatures finding a proper balance between legality and fine-grainedness—a balance that eluded the academic drafters of the Model Penal Code.

CONCLUSION

The theoretical analysis advanced here does not represent a consensus among punishment theorists. Far from it. The arguments of this paper are based on a novel theory of punishment that draws on a long-neglected moral theory. These arguments do, however, reflect a coherent account of the ends of punishment, of the centrally important concept of “culpability,” and of the underlying logic of that central concept. No matter its being a minority view, this analysis is an advance over the ad hoc illogic of commonplace punishment theory.