


1-1-2011

Modal Retributivism: A Theory of Sanctions for Attempts and Other Criminal Wrongs

Anthony M. Dillof

Wayne State University Law School

Follow this and additional works at: <https://scholarship.richmond.edu/lawreview>

 Part of the [Jurisprudence Commons](#), and the [Law and Philosophy Commons](#)

Recommended Citation

Anthony M. Dillof, *Modal Retributivism: A Theory of Sanctions for Attempts and Other Criminal Wrongs*, 45 U. Rich. L. Rev. 647 (2011).
Available at: <https://scholarship.richmond.edu/lawreview/vol45/iss2/7>

This Article is brought to you for free and open access by the Law School Journals at UR Scholarship Repository. It has been accepted for inclusion in University of Richmond Law Review by an authorized editor of UR Scholarship Repository. For more information, please contact scholarshiprepository@richmond.edu.

MODAL RETRIBUTIVISM: A THEORY OF SANCTIONS FOR ATTEMPTS AND OTHER CRIMINAL WRONGS

Anthony M. Dillof *

This article attempts to think systematically about what sanctions are deserved for a range of criminal offenses. The offenses considered include both consummate offenses (such as murder) and inchoate offenses (such as attempts), as well as offenses of negligence and crimes of passion. Unlike other theories of punishment, which tend to focus on justifications of the practice of punishment, or the “why” question, the theory presented here focuses on the amount of punishment, that is, the “how much” question: Specifically, “How much punishment, in terms of size and severity, is deserved for a given criminal offense?” The article attempts to answer this question for a variety of crimes in a unified, principled manner.

The article begins by considering and critiquing harm-based retributivism. Harm-based retributivism is a leading theory of punishment. According to it, the punishment an actor deserves depends on the harm she is culpably responsible for causing—the greater the harm, the greater the punishment deserved. Harm-based retributivism has normative appeal for many and is consistent with the great majority of our punishment practices.

The article then advances a novel criticism of harm-based retributivism. The problem with harm-based retributivism, it is argued, is that it cannot be extended in a principled manner to inchoate offenses, such as attempts. An alternative to harm-based retributivism is intent-based retributivism—the theory that punishment for both consummate and inchoate offenses should be based solely on subjective factors, such as the harm intended. In-

* Associate Professor, Wayne State University Law School. J.D., LL.M., Columbia University School of Law; A.B., Harvard University. I thank Steven Winter for his helpful comments and Maricanne Miller for her encouragement. For any wrongs herein, I of course am solely culpable.

tent-based retributivism, however, is inconsistent with a widespread intuition that "harm matters."

In response to these criticisms, the article presents an alternative to both harm-based and intent-based retributivism. The alternative theory presented is called, largely for lack of a better name, "modal retributivism." The essence of modal retributivism is that under it, the fact that harm results from wrongful conduct is not relevant to the size or severity of the sanction deserved, as it is under harm-based retributivism; rather, it is relevant to the sanction's mode as precatory ("should be imposed") or permissive ("may be imposed").

Rather than building the case for modal retributivism from the ground up, this article takes the existing components of retributive thought and reassembles them into a sounder structure. The cogency of the argument against harm-based retributivism and the appeal of modal retributivism will likely be strongest for those who allow reason, as opposed to intuition, a leading role in resolving moral issues.

I. HARM-BASED RETRIBUTIVISM

This part describes the basic theory of harm-based retributivism, sketches a defense of it, and shows how it might be elaborated to recommend specific sanctions for specific offenses.

A. *The Basics*

Retributivism is a normative theory of punishment. Its basic tenet is that people should be punished based on what they deserve.¹ Furthermore, the punishment people deserve is not a matter of the actual or predicted consequences of imposing the

1. There is no universally accepted definition of retributivism. See, e.g., John G. Cottingham, *Varieties of Retribution*, 29 PHIL. Q. 238 (1979) (describing nine varieties of retributivism). Some define retributivism strictly, holding that "[a] retributivist punishes because, and only because, the offender deserves it." Michael S. Moore, *The Moral Worth of Retribution*, in RESPONSIBILITY, CHARACTER, AND THE EMOTIONS 179, 179 (Ferdinand Schoeman ed., 1987). Others define it more loosely as the belief that "justice in punishment requires the features of a punishment to be shaped by reference to the features of the offense for which it is meted out." Hugo Adam Bedau, *Concessions to Retribution in Punishment*, in JUSTICE AND PUNISHMENT 51, 53 (J.B. Cederblom & William L. Blizek eds., 1977) (emphasis omitted).

punishment.² As such, retributivism is a nonconsequentialist theory of punishment.

Even though retributivists agree that punishment should be based on desert, they disagree on the proper analysis of desert. The prevalent view—what might be called the “received model”—is that the punishment an actor deserves is a function of (1) the gravity of the wrongdoing she commits; and (2) her responsibility, or culpability, for that wrongdoing.³ Among those who hold this view, perhaps the most fundamental point of disagreement is whether the harmful results of an actor’s wrongful conduct should be deemed part of the wrongdoing and thus relevant to desert. To take a well-worn example, imagine that *A*, intending to kill *B*, shoots a gun at *B*’s heart. Should the punishment that *A* deserves depend on whether the bullet strikes and kills *B* as intended, or misses *B* due to an unforeseeable gust of wind which blows the bullet off course? Proponents of harm-based retributivism answer yes: Assuming *B* is struck and killed, *A*’s wrongdoing is properly characterized as killing (or causing the death of) *B*; *A* is culpably responsible for that wrong. In the case of the bullet blown off course, the wrongdoing is merely *A*’s shooting at *B*. This wrongdoing is much less grave than killing *B*, and so *A* deserves less punishment.⁴ In contrast, intent-based retributivists hold the view that in both cases *A*’s wrongdoing is properly characterized as his shooting to kill *B*; it is irrelevant whether the

2. Bedau, *supra* note 1, at 53.

3. See GEORGE P. FLETCHER, *RETHINKING CRIMINAL LAW* § 6.6, at 455 (Oxford Univ. Press 2000) (1978); ROBERT NOZICK, *ANARCHY, STATE, AND UTOPIA* 59–63 (1974). A possible alternative view of desert is that punishment should be a function not of the wrongful disadvantage imposed upon an individual or society, but of the unfair advantage that the actor acquired by engaging in the prohibited conduct. *E.g.*, WOJCIECH SADURSKI, *GIVING DESERT ITS DUE: SOCIAL JUSTICE AND LEGAL THEORY* 225 (1985); GEORGE SHER, *DESERT* 77 (1987). Whether the wrongful disadvantage view and the unfair advantage view diverge in practice is debatable. See SHER, *supra*, at 81–82 (equating wrongdoer’s degree of benefit with his act’s degree of wrongness).

4. See JOSHUA DRESSLER, *UNDERSTANDING CRIMINAL LAW* § 27.04[B][3] (5th ed. 2009) (discussing the traditional view that criminal attempts are “lesser offense[s] than the consummated crime[s]”). In general, harm-based retributivists defend the view that results are relevant to desert. See R.A. DUFF, *CRIMINAL ATTEMPTS* 345 (1996); FLETCHER, *supra* note 3, § 6.6, at 482–83; Leo Katz, *Why the Successful Assassin Is More Wicked Than the Unsuccessful One*, 88 CAL. L. REV. 791 (2000); Michael S. Moore, *The Independent Moral Significance of Wrongdoing*, 5 J. CONTEMP. LEGAL ISSUES 237 (1994).

bullet killed *B* or harmed no one.⁵ In both cases, *A* deserves the same punishment.

So described, why should we be interested in harm-based retributivism? I do not intend to offer a full defense of the theory here. I seek merely to show it is *prima facie* of interest and plausible enough to justify the examination, critique, and revision that follows. Let us then consider the case for interest in harm-based retributivism.

First, there is the brute fact that there is an enormous amount of writing on retributivism, both in the legal and philosophical literature.⁶ If no longer trendy, retributivism is well-established and of continuing scholarly interest.⁷

Second, retributivism, in particular harm-based retributivism, has a claim to *prima facie* validity. Retributivism has been criticized for failing to provide a justification for the presumptively problematic practice of inflicting suffering upon fellow community members.⁸ Retributivism, it is said, should explain, not simply appeal to, the idea that the guilty deserve to suffer.⁹ Yet not all moral claims can be explained in terms of moral principles that are more certain. Sometimes moral bedrock has been hit and there are no more certain principles to which to appeal. Rather than quixotically seeking more and more fundamental principles

5. Those who would defend forms of intent-based retributivism argue punishment should not turn on matters of "luck," like whether a would-be victim escapes harm due to unforeseeable forces. See LARRY ALEXANDER & KIMBERLY KESSLER FERZAN, *CRIME AND CULPABILITY: A THEORY OF CRIMINAL LAW* 171, 192 (2009); Sanford H. Kadish, *Foreword: The Criminal Law and the Luck of the Draw*, 84 J. CRIM. L. & CRIMINOLOGY 679, 688 (1994); Stephen J. Morse, *Reason, Results, and Criminal Responsibility*, 2004 U. ILL. L. REV. 363, 379–81 (2004); Stephen J. Schulhofer, *Harm and Punishment: A Critique of Emphasis on the Results of Conduct in the Criminal Law*, 122 U. PA. L. REV. 1497, 1498 (1974).

6. See, e.g., ALEXANDER & FERZAN, *supra* note 5; JOEL FEINBERG, *DOING & DESERVING: ESSAYS IN THE THEORY OF RESPONSIBILITY* (1970); MICHAEL MOORE, *PLACING BLAME: A GENERAL THEORY OF THE CRIMINAL LAW* (1997); Jerry Cederblom & Cassia Spohn, *A Defense of Retributivism Against Criticisms of the Harm-for-Harm Principle*, 43 CRIM. L. BULL. 920, 920–24 (2007); DJ Galligan, *The Return to Retribution in Penal Theory*, in *CRIME, PROOF AND PUNISHMENT* 144 (1981); Herbert Morris, *Persons and Punishment*, 52 *MONIST* 475 (1968).

7. See generally Russell L. Christopher, *Deterring Retributivism: The Injustice of "Just" Punishment*, 96 NW. U. L. REV. 843, 845–47 (2002) (reviewing the rise of retributivism among legal scholars, philosophers, courts, and legislatures).

8. See R.A. Duff, *Penal Communications: Recent Work in the Philosophy of Punishment*, 20 *CRIME & JUST.* 1, 29 (1996).

9. *Id.* at 29; see also Cederblom & Spohn, *supra* note 6, at 922 (faulting Kantian retributivism in that it fails to explain why wrongdoers should be punished).

on which to base a moral claim, moral inquiry may sensibly begin by examining society's actual well-established moral views and working from there. In this light, harm-based retributivism does well. Empirical research reveals that harm-based retributivism reflects deeply rooted popular sentiment.¹⁰ Furthermore, the criminal law, society's most explicit statement about the relation between wrongdoing and punishment, largely reflects norms of harm-based retributivism. For example, in all jurisdictions, murder—the offense of unjustifiably and inexcusably taking the life of another—is punished more harshly than consummate offenses defined by lesser harms.¹¹ As a general matter, consummate offenses are punished more harshly than the corresponding attempted ones.¹² Although other explanations are surely possible, our society and criminal justice system have the look and feel of ones based on harm-based retributivism.

10. See, e.g., PAUL H. ROBINSON & JOHN M. DARLEY, *JUSTICE, LIABILITY, AND BLAME: COMMUNITY VIEWS AND THE CRIMINAL LAW* 14–28, 181–89 (1995) (reporting empirical studies supporting deeply held views among the community that resulting harm does matter); Dena M. Gromet & John M. Darley, *Punishment and Beyond: Achieving Justice Through the Satisfaction of Multiple Goals*, 43 *LAW & SOC'Y REV.* 1, 5, 12–14 (2009) (summarizing a variety of studies purporting to show public support for retributive justice, especially for morally grave results); John H. Mansfield, *Hart and Honoré, Causation in the Law—A Comment*, 17 *VAND. L. REV.* 487, 494 (1964) (concluding that “[t]he notion that there should be a difference in punishment [between unsuccessful attempts and consummate crimes] is deeply rooted in the popular conscience”); Janice Nadler & Mary R. Rose, *Victim Impact Testimony and the Psychology of Punishment*, 88 *CORNELL L. REV.* 419, 423 (2003) (discussing studies which indicate people generally base their punishment judgments on harm-based retributivism). *But see* Janine Natalga Clark, *The Limits of Retributive Justice: Findings of an Empirical Study in Bosnia and Hercegovina*, 7 *J. INT'L CRIM. JUST.* 463, 486–87 (2009) (concluding that victims of international human rights crimes may not be as supportive of retributive justice after participating in criminal trials of the accused).

11. The Supreme Court recently held the death penalty is an unconstitutional punishment for even an offense as serious as child rape. *Kennedy v. Louisiana*, 554 U.S. ___, ___, 128 S. Ct. 2641, 2646 (2008).

12. It has been said that “[t]he consensus of Western legal thought is to punish attempts less severely than consummated offenses.” FLETCHER, *supra* note 3, § 6.6, at 474; *see also* DRESSLER, *supra* note 4, §§ 27.02[A], [C]; JOHN KAPLAN, ROBERT WEISBERG & GUYORA BINDER, *CRIMINAL LAW* 663 (6th ed. 2008) (noting that most jurisdictions punish attempts less than consummate offenses). The Model Penal Code is arguably the exception that proves the rule. Except for capital offenses, the Model Penal Code punishes attempts at the same level as the consummate offense. MODEL PENAL CODE § 5.05(1) (1962). Jurisdictions that have revised their penal codes in light of the Model Penal Code, however, have consistently rejected this aspect of the Code. *See* KAPLAN, WEISBERG & BINDER, *supra*, at 663. Furthermore, the Model Penal Code sometimes takes results into account in determining a punishment. Reckless conduct that causes death results in a conviction for manslaughter. Reckless conduct that is harmless results only in a conviction of reckless endangerment, a lesser offense. *Compare* MODEL PENAL CODE § 210.3, *with id.* § 211.2. *See also* DUFF, *supra* note 4, at 117 & nn.6, 7 (discussing English and Canadian practice of imposing higher penalties for attempts than substantive offenses).

Third, on the merits, a strong case has been made for harm-based retributivism. Michael Moore, for example, has presented a well-reasoned contemporary defense of the doctrine.¹³ Moore himself puts little stock in foundationalism—the idea that the truth or falsity of harm-based retributivism is self-evident or that guidance regarding the truth or falsity of harm-based retributivism can be gleaned by appeal to more abstract, more abstruse, higher-order moral principles.¹⁴ Rather, Moore believes that a coherence approach in which a range of closely related moral phenomena are examined is the most fruitful way to shed light on harm-based retributivism's validity.¹⁵ Pursuing this approach, Moore argues that support for harm-based retributivism is to be found in (a) our intuitive response to cases of wrongdoing that involve, or do not involve, harm; (b) the greater intensity of our virtuous resentment of those who succeed in their wrongful attempts; (c) the greater feelings of guilt we experience when we succeed in causing harm compared to when we fail; (d) our forward-looking desire that we not only make the best choice given the facts available, but also that the choice turns out well; and (e) our related understanding that insofar as we are obligated to others, we experience the obligation as an obligation to accomplish, not to merely try.¹⁶ This network of related moral feelings about harm-based retributivism and the significance of consequences, though not logically entailing, are evidence that its appeal is not an anomaly, but is a highly integrated component of our overall moral makeup. If not uncontested, harm-based retributivism is then at least a vital moral theory.

B. *Harm-Based Retributivism Elaborated*

Much legal and philosophical scholarship has been devoted to the topic of justifying criminal punishment. This is the “why” question. Why do we punish for violations of the criminal law? To this question, a range of answers, both consequentialist and non-consequentialist, have been advanced. Relatively little, however,

13. See Moore, *supra* note 4.

14. *Id.* at 259–63.

15. *Id.* at 263–64.

16. *Id.* at 267–71. In a subsequent work, Moore defends these views against objections. MICHAEL S. MOORE, CAUSATION AND RESPONSIBILITY 23–33 (2009).

has been written on the “how much” question.¹⁷ How much should a person who violated the criminal law be punished?

Of course, the two questions are related. There should be as much punishment as needed to satisfy the purpose of punishment. This answer, however, fails to advance the ball. It is empty to say, as retributivists do, that criminals should be punished because they deserve to be and, in answer to the query “How much?” to respond “As much as they deserve.” If there was no more content to retributivism than that, it would lack sufficient substance to be evaluated. Some independent theory regarding the nature of desert would be needed. Even a more elaborate statement of desert, such as “a person deserves to be punished because she is culpably responsible for harming another,” would need to be supplemented before it could offer specific sanction recommendations which might be evaluated for soundness. If unable to offer such recommendations, a theory might be faulted on practical grounds for being incomplete and otiose, and on theoretical grounds for not being subject to challenge as being inconsistent with moral intuition.

In contrast, consider utilitarianism. A utilitarian believes that the right act is the one that maximizes social welfare.¹⁸ In the context of the criminal law, utilitarianism asserts that the correct amount of punishment maximizes social welfare by producing benefits to society, such as decreased criminal activity, that offset the loss to the offender, his family, and others due to his punishment.¹⁹ Of course, because of empirical uncertainties, the

17. See Andrew von Hirsch, *Proportionality in the Philosophy of Punishment: From “Why Punish?” to How Much?*, 1 CRIM. L.F. 259, 259 (1990) (“The philosophical writings have paid comparatively little attention to the criteria for distributing punishments—particularly to the criteria for deciding how much to penalize convicted offenders”). Antony Duff, in the most sustained and sophisticated contemporary analysis of criminal attempts, offers a draft attempt statute and argues that attempts should be punished less than substantive offenses. See DUFF, *supra* note 4, at 352–54, 398. Duff, however, offers little guidance as to the level at which attempts should be punished relative to substantive offenses. *Id.* at 398–99. Likewise Larry Alexander and Kimberly Ferzan, in their comprehensive treatment of criminal law based on the meting out of retributive justice—which includes an extensive discussion of sentencing considerations—fail to suggest any absolute levels of appropriate punishment. ALEXANDER & FERZAN, *supra* note 5.

18. DRESSLER, *supra* note 4, § 2.03[A].

19. See JEREMY BENTHAM, AN INTRODUCTION TO THE PRINCIPLES OF MORALS AND LEGISLATION 158, 165 (J.H. Burns & H.L.A. Hart eds., Clarendon Press 1996) (1789); Michael D. Boyles, *Introduction to CONTEMPORARY UTILITARIANISM* 1, 1, 4 (Michael D. Boyles ed., 1968). See generally Guyora Binder & Nicholas J. Smith, *Framed: Utilitarianism and Punishment of the Innocent*, 32 RUTGERS L.J. 115, 140–54 (2000) (offering a sophisticated

correct sanction level for any given offense may not be ascertainable to any degree of precision or certainty. But at least utilitarianism claims to be a complete theory of punishment on the ground that, if sufficient facts were known, the answer to the question "How much should an actor be punished?" could be had. As a normative theory, it needs no supplementation.

It is not clear how retributivism (the theory that punishment is justified because wrongdoers deserve to suffer) is to be supplemented to meet what might be called the "completeness criterion." There is a general consensus spanning the harm-based retributivist/intent-based retributivist divide that justice, at a minimum, requires proportionality between the gravity of the wrongdoing and the magnitude of the sanction; the greater the wrongdoing, the greater the sanction.²⁰ The proportionality principle critically supplements the idea of desert, giving it some content. Can more be said? Proportionality implies some ratio between the amount of wrongdoing and the magnitude of the sanction. What should the ratio be between the amount of wrongdoing and the amount of sanction? 3:1? 1:2? 5:7? 13:11? Or should the ratio be variable, for example, sometimes 5:7 and sometimes 13:11, reflecting a more complicated relation between wrongdoing and sanction? Or might all sanctions within a certain range, say from 2:5 to 7:6 (wrongdoing to sanction), be equally acceptable?

The most common answer given by retributivists is that the ratio of harm to punishment in the case of intentionally caused harms should be 1:1.²¹ An extreme version of this position is known as "*lex talionis*." *Lex talionis* is a theory of punishment with roots in the Code of Hammurabi and the Old Testament.²² It

account of Bentham's utilitarian views applied to punishments for crime).

20. DRESSLER, *supra* note 4, § 2.03[B][1].

21. See JEAN HAMPTON, *The Retributive Idea*, in FORGIVENESS AND MERCY 111, 127 (Jeffrie G. Murphy & Jean Hampton eds., 1988) (noting that "retributivists typically endorse the *lex talionis* as a punishment formula"); IMMANUEL KANT, *METAPHYSICAL ELEMENTS OF JUSTICE* 140 (John Ladd trans., 2d ed. Hackett Pub'g Co. 1999) (1797) ("[E]veryone [should] duly receive what his actions are worth."); Ernest van den Haag, *The Death Penalty Once More*, 18 U.C. DAVIS L. REV. 957, 971 (1985) (arguing that the death penalty is a "congru[ent]" punishment to murder). In the context of developing a theory about the upper limit of just punishment, Lawrence Crocker criticizes proportionality as lacking determinacy and insists that "[t]he only acceptable position is the identity of the upper limit of the penalty with the seriousness of the offense." Lawrence Crocker, *The Upper Limit of Just Punishment*, 41 EMORY L.J. 1059, 1104 (1992).

22. See 2 THE BABYLONIAN LAWS §§ 196, 197, 200, 210, 229, 230, at 77-83 (G.R. Driv-

is often encapsulated in the phrase “[an] eye for [an] eye.”²³ According to *lex talionis*, the wrong done by an actor should be visited on the actor.²⁴ Put somewhat less paradoxically,²⁵ the sanction imposed on the wrongdoer should be the same type of disadvantage as that suffered by the victim. Raping the rapist or torturing the torturer, however, is usually considered unacceptable even if it does not violate the rights of the offender.²⁶ Such punishments plausibly degrade the person inflicting them. Accordingly, sounder versions of *lex talionis* allow for similarity of sanction and wrongdoing on an abstract level that does not entail conduct that is otherwise objectionable.²⁷ In the case where another is intentionally blinded, for example, the wrongdoer, rather than being blinded, might be made to suffer an equally serious deprivation of stimulus (solitary confinement); or more abstractly, an equally serious physical impairment (shackles or hard labor); or, even more abstractly, an equally serious freedom limitation of some type (lengthy prison term).²⁸ At whatever level of abstraction the sanction resembles the wrongdoing, however, there remains an equivalence at the most abstract level: severity, or amount, of disadvantage, however instantiated.²⁹ Borrowing

er & John C. Miles eds. & trans., 1955); *Exodus* 21:24–25 (New King James); see also *Genesis* 9:6 (New King James); *Leviticus* 24:19–20 (New King James); *Deuteronomy* 19:21 (New King James).

23. *Exodus* 21:24 (New King James).

24. E.g., BLACK’S LAW DICTIONARY 996 (9th ed. 2009) (defining *lex talionis* as “[t]he law of retaliation, under which punishment should be in kind—an eye for an eye, a tooth for a tooth, and so on—but no more”).

25. The paradox is that the wrongdoer not only harms her victim, but, critically, wrongs her victim. But harming the wrongdoer does not wrong the wrongdoer. She deserves it. So, by definition, any harm visited upon her lacks the critical moral feature of the harm she visited on another.

26. Similarity of type between harm caused and sanction is not generally practical. As Blackstone recognized, “[t]heft cannot be punished by theft, defamation by defamation, forgery by forgery, adultery by adultery.” 4 WILLIAM BLACKSTONE, COMMENTARIES *13; see JOHN BRAITHWAITE & PHILIP PETTIT, NOT JUST DESERTS: A REPUBLICAN THEORY OF CRIMINAL JUSTICE 148 (1990) (arguing *lex talionis* is flawed because it recommends punishments for offenses such as rape and bribery).

27. See Jeremy Waldron, *Lex Talionis*, 34 ARIZ. L. REV 25, 25, 33–37, 43–45 (1992) (arguing that a sounder version of *lex talionis* only requires similarity of the essential wrongness-making features of the offense); see also Michael Davis, *Harm and Retribution*, 15 PHIL. & PUB. AFF. 236, 238 (1986) (defining *lex talionis* as a principle of equal suffering, not equal types of suffering).

28. Some societies, however, adhere to a literal interpretation of *lex talionis*. See, e.g., Robert Tait, *Eye for an Eye: Iranian Man Sentenced to be Blinded for Acid Attack*, GUARDIAN, Nov. 28, 2008, <http://www.guardian.co.uk/world/2008/nov/28/iran-acid-attack-sharia-law>.

29. In this respect, retributive justice tracks corrective justice—a form of justice that

from Kant, I shall refer to this idea of equal magnitude, or 1:1 ratio between harm and sanction, as the principle of equality.³⁰

I believe that harm-based retributivism, when supplemented with the principle of equality, satisfies the completeness criterion—or at least comes reasonably close. Admittedly, further theoretical work remains. Some theory is needed regarding the level of abstractness at which the sanction must mirror the harm. Possibly this theory is along the lines of “as much as possible given practical limits, X, Y, and Z.” Likewise, some theory is needed to convert harm caused to years of incarceration or other sanction. Such a theory might provide that a given harm and a term of incarceration were equivalent if the offender, the victim, or perhaps a reasonable person, would be indifferent between them.³¹ Other questions still remain; for example, if *A* kills *B*, does executing *A* impose a sanction equal to the harm caused, or must *A*'s life be shortened by the (possibly brief) amount of time *B*'s life was shortened? Nevertheless, because we can envision the lines along which questions might be answered, I judge harm-based retributivism supplemented by the equality principle to satisfy the completeness criterion.

II. A CRITIQUE OF HARM-BASED RETRIBUTIVISM

This part presents an argument against harm-based retributivism. The strength and interest of the argument is that it is an internal criticism of harm-based retributivism. Rather than seek-

I believe we have stronger and better defined intuitions about. Where a person tortiously injures another or destroys another's property, the ideally just response from the perspective of corrective justice would be for the actor to heal the victim's injuries or recreate the destroyed property—to literally restore the status quo ante. Where this is impractical, monetary compensation is regarded as fair insofar as the amount might restore the victim to his ex ante welfare level.

30. In support of the principle of equality, Kant wrote:

What kind and what degree of punishment does public legal justice adopt as its principle and standard? None other than the principle of equality (illustrated by the pointer on the scales of justice), that is, the principle of not treating one side more favorably than the other. Accordingly, any undeserved evil that you inflict on someone else among the people is one that you do to yourself. If you vilify him, you vilify yourself; if you steal from him, you steal from yourself; if you kill him, you kill yourself.

KANT, *supra* note 21, at 138. Crocker adopts Kant's principle of equality as setting the upper limit of just punishment. See Crocker, *supra* note 21, at 1061, 1065, 1067–79 (defending “reciprocity” as setting the upper limit of punishment).

31. See Cederblom & Spohn, *supra* note 6, at 924–25 (recommending such an approach).

ing to dislodge the proponent of harm-based retributivism from her position based on arguments from premises that for many have less intuitive appeal than harm-based retributivism itself, the criticism here attempts to show the difficulty inherent in working through the intuitions that underlie harm-based retributivism. I will show that harm-based retributivism, while perhaps invulnerable from attack by opposing moral intuitions, may be critiqued from what I shall describe as the vantage point of reason.

A. *Reason, Intuition, and Morality*

Moral questions are neither questions of empirical science nor mathematics. They cannot be settled by experiment or formal proof. Furthermore, argumentation has a limited place. When it comes to moral arguments, it is too easy to admit premises, but deny conclusions when the alleged connection between claims is analogy rather than logic. Rather than experiment, proof, or argument, moral decisionmaking often relies heavily on direct appeal to moral intuitions. Reliance on moral intuitions is especially important when it comes to fundamental issues, such as whether harm—or intent-based—retributivism is right. Sadly, moral intuitions are not distinct, clearly labeled psychological entities with special claims to truth or authority. They are just your/my/our feelings about what is sound, decent, sensible, and just. Such feelings are often an immediate emotional reaction to actual or imagined situations in which the issue directly arises or is indirectly implicated. Sometimes we can say no more than, “Upon reflection, it feels right/wrong to me.”³²

Conceding the importance of moral intuitions, there is a distinct place in moral decisionmaking for reason. To return to the question of the proper proportion between harm and punishment, the principle that the wrongdoing and sanction should be equivalent (an eye for an eye), rather than the principle that, say, the wrongdoing should be 7/9 or 9/7 of the sanction, is supported by

32. According to Moore, “We need our emotions to know about the injustice of racial discrimination, the unfairness of depriving another of a favorite possession, the immorality of punishing the innocent. Our emotions are our main heuristic guide to finding out what is morally right.” Moore, *supra* note 1, at 189; *see also* Moore, *supra* note 4, at 262 (“It is also difficult to think of some more general, more basic principle with which to justify [harm-based retributivism]—‘causation just matters’ seem[s] to be about the only argument one can give.”).

reason rather than moral intuition. Seven-ninths or 9/7 of the wrongdoing seems unacceptably arbitrary.³³ It does not seem a plausible candidate for a basic feature of morality. The rules of morality should be principled and nonarbitrary. What counts as principled and nonarbitrary is a matter of reason.

Furthermore, the appeal of the equality principle appears not to derive merely from unmediated moral reactions or an intuitive sense of justice. Equivalence between wrongdoing and sanction is difficult to precisely calibrate. For example, assume that our considered and settled moral intuition is that fifteen years' incarceration is appropriate for a serious instance of kidnapping. What general moral principle does this moral intuition datum support? Given the difficulty of establishing that fifteen years is precisely equivalent to the harm caused by the kidnapping, this intuition seems as consistent with the principle that the sanction should be equivalent to 49/50 of the harm as it is with the principle that harm and sanction should be equal. To the extent that the latter seems a more likely true moral principle than the former, its appeal is not based on a generalization from individual instances of moral reaction (which equally well support very close but distinct principles), but the appeal of an independent principle of reason.

The idea that reason is a faculty for discovering moral truth, and that moral truths have a simple and elegant formulation, not unlike the laws of nature, may strike the reader as discredited Platonism. This article takes no position on the nature of moral truth, whether it exists independent of the human mind, how moral truth claims compare to truth claims about the physical world, or other philosophical issues. Regardless of how these enormously difficult issues are to be resolved, the fact remains that when thinking about questions of justice, we give great deference to considerations of nonarbitrariness, just as we do in other areas. Consider, for example, corrective justice. As retributive justice is the norm that animates criminal law, corrective justice is the norm that animates tort law. Under tort, if *D* wrongfully causes harm to *P*, *D* is liable and *P* is entitled to damages. But if, for example, *D* is responsible for *P*'s broken arm, what damages

33. Kant appears to have championed such a position when arguing that *lex talionis* is the only principle that can "specify definitely the quality and the quantity of punishment," as "all other principles are fluctuating and unsuited." IMMANUEL KANT, THE METAPHYSICS OF MORALS 106 (Mary Gregor ed. & trans., Cambridge University Press 1996) (1797).

does *P* deserve? A payment equivalent to 7/9 or 9/7 of the harm suffered? Because pain and suffering cannot be directly translated into dollars and cents, again it makes little sense to claim that the principle that the damages should be equivalent to the harm is simply a generalization of visceral ideas about what is just in specific cases, or to approach the question of the proper amount of damage through intuitions unmediated by the general principle that the amount of compensation should be equivalent to the harm. This principle, as opposed to the principle that a tort victim should receive compensation equivalent to 7/9 or 9/7 of the harm suffered, appeals to reason and seems a candidate for a fundamental principle of justice.

Few this side of Immanuel Kant believe that reason alone can decide moral controversies, such as the perennial debate between harm-based and intent-based retributivists. But the arbitrariness of a moral theory can be a strong objection to that moral theory as, for example, arbitrariness was a strong objection to the theory that the proportion of harm to sanction should be 7:9. Reason thus not only can supplement moral intuitions at the margins in specific cases, it can at least sometimes, if not identify the correct moral theory, disqualify incorrect ones.

B. *The Problem of Attempts*

Having hopefully laid the necessary groundwork, I now begin the substantive part of the argument.

1. The Basic Problem

Consider the following possible objection to harm-based retributivism:

According to the most plausible version of harm-based retributivism, the amount of harm culpably caused should equal the severity of the sanction (however magnitude of harms and sanctions should be measured). According to harm-based retributivism, what should the sanction be for intended harm that does not result in harm? This is the problem of punishing attempts. Obviously, a sanction equal in severity to the harm intended cannot be the answer since the defining feature of harm-based retributivism is that harm matters to desert and a sanction of equal severity is what would be imposed had the harm occurred. Alternatively, no sanction at all, though principled, is not an intuitively acceptable position. It does not seem right that a person who tries to cause perhaps a grievous harm

should always go scot-free. No harm-based retributivism proponent has espoused the position that those who attempt to cause harm but “luckily” fail should thereby be free from all punishment. Finally, any sanction between one equal to the harm and no sanction—for example, a sanction equivalent to 7/9 of the harm intended—seems unacceptably arbitrary. The punishment people deserve in the absence of caused harm seems like a basic question of morality. Yet a 2/9 “failure discount” for cases where no harm is caused simply does not seem like a plausible candidate for a basic principle of morality. Other discounts seem equally arbitrary.³⁴ Harm-based retributivism is stuck when it comes to punishing attempts.

How should this objection be met?

2. Some False Starts

The amount of punishment for an attempt should not be arbitrary. Yet any fixed ratio between the sanction for attempt and the substantive offense seems, from a moral perspective, arbitrary. Accordingly, an attractive approach to fixing the level of sanctions for attempts is varying the failure discount based on some morally relevant feature of the attempt. The criminal justice system might then set that level of punishment for the attempt based on the factor itself or—where it is too difficult to determine the degree to which the factor is present—the level of punishment for the attempt might be set based on some average presence of the factor. Below four arguably morally relevant features of attempts are considered.

a. Proximity to Completeness

Criminal law distinguishes between mere preparation, for which there is not attempt liability, and perpetration, for which there is.³⁵ Traditionally, common law jurisdictions would subject to attempt liability only those actors who were sufficiently close to completing the offense they intended to commit.³⁶ Proximity to

34. Some criminal codes establish that the minimum and maximum penalties for attempts are 1/2 the minimum and maximum of the consummate offense. *See, e.g.*, GA. CODE ANN. § 16-4-6(b) (2007). One-half is less arbitrary than 7/8, but still seems unacceptably arbitrary because 1/2 is not linked to any general moral theory, nor is it a familiar constant in the moral universe.

35. *See, e.g.*, DRESSLER, *supra* note 4, §§ 27.02[B], 27.06[A].

36. *See id.* § 27.02[D]; *see also* WAYNE R. LAFAVE, CRIMINAL LAW § 11.2 (4th ed. 2003).

completeness is arguably morally relevant. Thus, one idea would be for the punishment for attempt to be based on a sliding scale, where the closer the wrongdoing to completion, the greater the sanction. If an actor were halfway to completing an offense that would result in harm H , she might be subject to a sanction of $H/2$. The problem with this approach is that it is not clear how such an approach might be implemented on a principled basis. What does it mean to be halfway to completing an offense? There seems no plausible metric available for measuring closeness to completing an offense for moral purposes. Miles and hours seem a wholly arbitrary measure of closeness to completion.

b. Intensity of Desire

Criminal law also requires that an actor, in order to be liable for an attempt, intend to commit a substantive offense.³⁷ It seems clear that the moral justification for punishing a person for an attempt is closely tied to the fact that the actor's conduct was intended to harm another. Intent is generally bivalent: a result is either intended or it is unintended. Yet in some contexts, it is coherent to speak of degrees of intent. One person may intend to reach a goal more than another in the sense that one actor has a stronger desire to achieve the goal than another actor. A potential theory of punishment for attempts is that the magnitude of the sanction that an actor deserves for an attempt should be a function of the strength of the actor's desire that the attempt succeed—the greater the strength, the greater the deserved punishment. This would allow the actor's punishment to vary with the magnitude of a morally critical feature of the attempt. It is at least plausible that those with a greater commitment to wrongdoing deserve greater punishment. Whatever its appeal, however, the trouble with such an approach is that it does not lead to assigning sanctions for punishments in a nonarbitrary way. Suppose A and B independently attempt to kill V and fail. Assume that A 's desire to kill V is twice B 's desire in the sense that A would invest up to \$2000 to kill V and B would invest only up to \$1000. What sanctions should they receive for attempted murder? Even if A should be punished twice as much as B , critically we are given no guidance on absolute punishment level.

37. See, e.g., DRESSLER, *supra* note 4, §§ 27.02[B], [D], 27.05[A].

c. Secondary Harms

A third possible determinant of the sanction level for attempts might be the “secondary harms” associated with the attempt. The primary harm associated with an offense is the harm that definitionally must occur for the offense to exist—for example, in the case of murder, the death of another person. Even in cases of failed attempts, however, society is usually harmed to some degree. Sometimes a would-be victim is aware of or learns of the attempt and suffers fear, anxiety, or other psychological harm. Sometimes members of the community learn of the attempt and suffer some sort of vicarious emotional harm or they expend resources to protect themselves against similar crimes. Other times other persons are emboldened to act in similar ways. These sorts of harms are secondary. The harm-based retributivism proponent might fasten onto such harms as providing both the moral basis for punishing failed attempts and the measure of sanction as well.³⁸ A nonarbitrary theory of the punishment of attempts is that the sanction should be equal in severity to the secondary harm caused.

The difficulty with this theory is that it is far from clear that it justifies punishment levels at anywhere near the level that intuitively is thought proper. Often a failed attempt will result in no or minimal harm to society. Furthermore, even when a secondary harm occurs, it is usually unintended, frequently unforeseen, and sometimes unforeseeable. Accordingly, the actor probably deserves substantially less punishment for causing these harms than if the harms were intended. Thus, looking to these harms as a basis of punishment does not seem to get the harm-based retributivism proponent where she wants to go.

38. See Lawrence C. Becker, *Criminal Attempt and the Theory of the Law of Crimes*, 3 PHIL. & PUB. AFF. 262, 271 (1974) (arguing that attempts should be punished as severely as consummate crimes because both cause the same amount of social harm); see also DUFF, *supra* note 4, at 134 (identifying the prevention of secondary harms as a justification for attempt liability); JEROME HALL, *GENERAL PRINCIPLES OF CRIMINAL LAW* 218–19 (2d ed. 1960) (identifying the harm of inchoate crimes with “apprehension aroused” and increased risk of other crimes).

d. Unfair Advantage

Michael Davis has developed an elaborate theory for determining the just punishment for a criminal offense.³⁹ Davis's theory focuses not on the harm done by the criminal act, as does *lex talionis*, but on the unfair advantage the offender secures by engaging in the act.⁴⁰ Davis argues that unfair advantage should be understood by reference to a hypothetical market in which licenses to commit crimes were sold to the highest bidder; the price of a license corresponds to the unfair advantage secured by a criminal.⁴¹

Davis argues his theory yields an answer to the question of how attempts should be punished compared to consummate offenses.⁴² According to Davis, attempts deserve to be punished at a significant level, yet less than consummate offenses, since licenses to commit attempts will bear a lower market price than licenses to commit consummate offenses.⁴³ Davis's desert theory, however, is only a theory of the ordinal ranking of offenses.⁴⁴ Consequently, it cannot be used to determine a general relationship between the absolute severity of punishment for attempts and consummate offenses, the issue under examination here.

Furthermore, Davis's substantive argument that licenses for attempts will sell at a lower price than licenses for consummate offenses is flawed because it assumes that licenses for consummate offenses are also licenses for attempts and so logically have wider applicability. In fact, many criminal codes contain provisions that distinguish the requirements, and therefore the penalties, for consummate offenses and attempts.⁴⁵ Based on these

39. See MICHAEL DAVIS, *How to Make the Punishment Fit the Crime*, in *TO MAKE THE PUNISHMENT FIT THE CRIME* 69 (1992).

40. See Davis, *supra* note 27, at 240.

41. Michael Davis, *How to Make the Punishment Fit the Crime*, 93 *ETHICS* 726, 742–46 (1983).

42. *Id.* at 744 n.18.

43. See *id.*

44. Michael Davis, *Criminal Desert and Unfair Advantage: What's the Connection?*, 12 *L. & PHIL.* 133, 140 (1993).

45. See, e.g., 18 U.S.C. § 2332(a)–(b) (2006) (distinguishing between penalties for the consummated offense of murder of a United States national outside of the United States from an attempt of such murder); MICH. COMP. LAWS ANN. § 750.92 (West 2004) (defining attempts to exclude the successful commission of an offense). Compare VA. CODE ANN. § 18.2-32 (Repl. Vol. 2009) (defining punishment for first and second degree murder), with

provisions, attempts and consummate offenses function as non-overlapping crimes. Essentially, attempts are unsuccessful efforts to commit crimes and consummate offenses are successful efforts. A license to engage in a consummate offense would not provide the right to engage in an attempt any more than a license to commit burglary would provide a right to commit bribery. Thus, for an offense where failure was more common than success, a license for an attempt, being more useful, might command a higher price than a license for the consummate offense, implying a higher penalty is deserved. Such a result, however, is wildly out of conformity with actual practice and intuitions.

3. The Risk Response to the Basic Problem

a. The Basic Idea

One natural move open to harm-based retributivists at this juncture is to look to the riskiness of conduct. The riskiness of conduct might be thought to ground its wrongfulness analogously to the way that the harmfulness of conduct may ground its wrongfulness.⁴⁶ Risk, as I shall use the term, is an epistemic concept measuring the subjective likelihood that a harm will occur.⁴⁷ It is plausible to think that we have a moral duty not to engage in conduct which we believe risks harming another, and that this duty is violated when one engages in such conduct, even in cases where the risk does not materialize as harm. For example, one can imagine a hypothetical story of people contracting not to engage in conduct the actor perceives as risking another's interests for two reasons: first, such conduct often results in harm and so is not favored; second, risk-creating conduct (as opposed to conduct actually resulting in harm) is conduct people can directly control whether they are engaging in it or not. Such a hypotheti-

id. § 18.2-26 (defining punishment for attempted noncapital felonies). Likewise, many criminal codes prohibit in cases of a completed offense conviction for attempting the offense as well. See, e.g., MODEL PENAL CODE § 1.07(1)(b) (1962).

46. The idea that risk imposition is analogous to harm creation has been explored by Claire Finklestein, who defends the claim that exposure to risk is itself a harm. Claire Finklestein, *Is Risk a Harm?*, 151 U. PA. L. REV. 963, 973 (2003). According to Finklestein, "[A]gents have a legitimate interest in avoiding unwanted risks." *Id.* at 966.

47. *Id.* at 973 ("[T]he subjective, or 'epistemic,' interpretation maintains that there is no such thing as an objective measure of probability. . . . [A]lthough an agent's degree of belief will be based on real observations he can make, likelihoods cannot be a matter of objective facts.").

cal agreement might be thought to inform our actual moral obligations.⁴⁸ Acting inconsistently with such a hypothetical agreement could be considered an instance of wrongdoing, perhaps not as severe as actually causing harm, but wrongdoing nonetheless.

Once engaging in risky conduct is identified as wrongdoing, it is but a short step to the position that the severity of the sanction should be a function of the level of the risk and the magnitude of the harm risked. Multiplication of the risk R by the amount of risked harm H seems the simplest and most intuitively appealing way of reckoning the sanction based on risk and harm. We frequently multiply likelihood by payoff to assess the value of a course of conduct, for example being indifferent between a .6 chance of a gain in welfare W and a .3 chance of a gain in $2W$.⁴⁹ Thus, where an actor intends to cause harm H and engages in conduct that the actor believes has a .8 chance of success, the actor might be subjected to a sanction of magnitude $.8 \times H$.⁵⁰ Rational actors who engage in conduct to achieve some harm (an attempt) believe that the conduct has some chance of success—otherwise they would not engage in it to achieve their end. They also believe the chance of success is less than 100% because they know the future is always contingent and can never be predicted with certainty. Thus, they assess the chances of their success as greater than 0% and less than 100%. If the sanction for trying to cause a harm H through conduct with a subjective likelihood of success R was punished by a sanction of magnitude $R \times H$, attempts would always be punished, and would always be punished

48. Rawls's famous argument for the legitimacy of democratic institutional structures based on their selection behind a "veil of ignorance" is an example of an appeal to hypothetical agreements. See JOHN RAWLS, *A THEORY OF JUSTICE* 118–23 (Harv. Univ. Press rev. ed. 1999); see also THOMAS HOBBS, *LEVIATHAN* 128–29 (A.P. Martinich ed., Broadview Press 2005) (1651) (emphasis omitted) ("I authorize and give up my right of governing myself to this man, or to this assembly of men, on this condition: that thou give up thy right to him, and authorize all his actions in like manner."); PHILIP PITT, *REPUBLICANISM: A THEORY OF FREEDOM AND GOVERNMENT* 63 (1997) (arguing that the theory of social contract is still valid, but that the absence of rebellion against a social contract, not consent of the ruled, is the only legitimacy that such a contract has).

49. See, e.g., Nicolas Gisiger, *Risk-Neutral Probabilities Explained* 11 (Working Paper, 2009), available at <http://ssrn.com/abstract=1395390>.

50. An analogous approach is taken by Crocker. He takes the position that an act can be wrongful by virtue of the objective risk it imposes on society. On this assumption, he argues that "an offense creating a risk of $1/n$ of concrete harm h imposes a harm of h/n ," which sets the upper limit of the punishment that may be justly imposed on the actor. Crocker, *supra* note 21, at 1084. Likewise, Alexander and Ferzan advocate discounting expected harm by probability in this manner when calculating deserved punishment. ALEXANDER & FERZAN, *supra* note 5, at 282–83.

less than in cases where the intended harm was actually achieved. Although what I shall call the “harm/risk” theory of punishment does not have the simplicity of the original harm theory, it has the virtue of providing a relatively principled, intuitively acceptable answer to the question of what the punishment differential should be between wrongful acts that result in harm and those that do not. The differential should strictly depend on the riskiness of the acts that do not produce harm. It also, of course, yields the result, consistent with the standard version of harm-based retributivism, that unsuccessful attempts are punished less than consummate offenses because $R \times H$ will be less than H .

b. Problems with the Risk Response

Despite its ability to yield specific punishment recommendations, the harm/risk theory sketched above is unsatisfactory. One problem appears in the following relatively specialized case. Compare $A1$ and $A2$. $A1$ intends to kill $B1$, engages in conduct that subjectively has a 100% chance of causing $B1$'s death, and so causes $B1$'s death. $A2$ intends to kill $B2$, engages in conduct that subjectively has a 100% chance of causing $B2$'s death, but fails to cause $B2$'s death. Under the harm/risk-based theory described above, the sanctions for $A1$ and $A2$ would be the same—a sanction equivalent to H . Such a result, however, fails to reflect the core intuition of harm-based retributivism: that results matter. Thus harm/risk-based retributivism does not seem to be a fully consistent elaboration of harm-based retributivism.

I do not consider the problem above to be so serious that the harm-based retributivist must reject the harm/risk theory and be left without a theory of punishment for attempts. First, cases where the risk of harm is 100% are rare. It is not clear that it is ever rational to assign a 100% chance to any future event occurring. Second, cases where the risk of harm is perceived as 100% and the harm is not realized are even rarer. In most cases where a person is certain that something will happen, it in fact happens. Perhaps moral intuitions simply break down in such exceptional cases and the lack of a punishment differential between $A1$ and $A2$ can be swallowed by the harm-based retributivist as just an isolated exception to his theory.

There is, however, a more general problem for the harm-based retributivist who wishes to look to risk as the basis of a principled theory of punishment for attempts. The flaw is that the theory cannot be consistently elaborated in a way to handle other categories of putatively punishable acts. For example, what sanction is appropriate in cases of recklessly caused and recklessly risked harm? A harm is recklessly caused when an actor who does not intend to cause the harm knowingly creates a risk of the harm and the risk is realized.⁵¹ A harm is recklessly risked when the risk is not realized. The offense of reckless endangerment is an example of an offense based on recklessly risked harm.⁵²

Consider these four cases where an actor *A* engages in conduct that risks causing a harm *H* (for example, death) to *B*:

	Harm <i>H</i>	No Harm
Harm <i>H</i> intended; <i>R</i> risk of <i>H</i>	Case 1 Example: Murder Sanction severity: <i>H</i>	Case 2 Example: Attempted murder Sanction severity: $R \times H$
Harm <i>H</i> unintended; <i>R</i> risk of <i>H</i>	Case 3 Example: Reckless homicide Sanction severity: ?	Case 4 Example: Reckless endangerment Sanction severity: ?

The sanction in Case 1 is based on the theory of equality. The sanction in Case 2 is based on the harm/risk theory. What should the sanction be for Case 4? Per hypothesis, in both Case 2 and Case 4, the same level of risk is created. The difference between the actors in Case 2 and Case 4 is that they have a different attitude toward the perceived riskiness of their conduct. The actor in

51. Model Penal Code reckless homicide and common law depraved heart killings are examples of offenses of recklessly caused harm. See MODEL PENAL CODE § 210.3(1) (1962).

52. See *id.* § 211.2.

Case 2 desires it; the actor in Case 4 does not. Should this bear on their punishment?

In my view, no. In Case 4, *A knows of* the risk of harm to *B* that she creates. For example, *A* drives down a road at a high rate of speed, aware that in doing so there is a risk *R* that a pedestrian will be killed, but has no desire to kill, and has no purpose in creating the risk. (*A* would be as happy if there were no one in the area who was endangered by his speeding.) In contrast, in Case 2, *A purposely* creates the risk to *B*, hoping the risk will be realized. For example, *A* drives down the road at a high rate of speed in order to create a risk that *B*, her enemy, who is standing on the sidewalk, will be killed. The criminal law has traditionally treated purpose and knowledge as functionally equivalent states.⁵³ For example, engaging in conduct with knowledge that another's death will result and engaging in conduct with the purpose that another's death will result are both murder when death in fact results.⁵⁴ Treating purpose and knowledge equivalently is sensible because those who act with purpose and those who act with knowledge have a very high degree of responsibility toward their intended goal and known object. They both fully accept what will occur. Furthermore, consistent with the themes of this article, there appears no principled differential in sanction between those who purposely cause harm *H* and those who knowingly do so. If an actor who intentionally causes harm *H* should suffer a sanction of *H* severity, what is the severity of the harm that should be suffered by an actor who merely knows she will cause *H*? $(8/9) \times H?$ $(7/8) \times H?$ I have argued that creating risk may be considered a form of wrongdoing analogous to causing harm. If purposely causing harm *H* and knowingly causing harm *H* are to be treated the same, then it seems to follow that purposely engaging in risky conduct *R* and knowingly engaging in risky conduct *R* should be treated the same. Hence Case 4, like Case 2, calls for a sanction of $R \times H$.⁵⁵

53. See LAFAYE, *supra* note 36, § 5.2; see also MODEL PENAL CODE § 2.02 cmt. 2 (Official Draft and Revised Comments 1985) (explaining that "the Code draws a narrow distinction between acting purposely and knowingly, [representing] one of the elements of ambiguity in legal usage at the word 'intent'").

54. MODEL PENAL CODE § 210.2 cmt. 3 (Official Draft and Revised Comments 1980). Under the Model Penal Code, engaging in conduct with the purpose of causing an unlawful result and engaging in the conduct with the belief it "will cause" a prohibited result are both treated as attempts. *Id.* § 5.01(1)(a)-(b) (1962).

55. Alexander and Ferzan reach a similar conclusion in the context of developing a

But this conclusion leads to a problem. If the actors in both Case 2 and Case 4 deserve a sanction $R \times H$ (because they are equally responsible for the risk they create), what should the sanction be for Case 3, which is like Case 4, but for the occurrence of the risked harm?

	Harm H	No Harm
Harm H intended; R risk of H	Case 1 Example: Murder Sanction severity: H	Case 2 Example: Attempted murder Sanction severity: $R \times H$
Harm H unintended; R risk of H	Case 3 Example: Reckless homicide Sanction severity: ?	Case 4 Example: Reckless endangerment Sanction severity: $R \times H$

On one hand, the proponent of harm-based retributivism seems committed to a greater punishment than $R \times H$ in Case 3 because $R \times H$ is the sanction in Case 4. The harm-based retributivist believes that resultant harm is relevant to punishment. On the other hand, a sanction equivalent to H seems counterintuitive. Case 3 should be punished less harshly than Case 1 because in Case 3, A 's culpability with respect to the harm seems much less than in Case 1. The Model Penal Code generally punishes reckless homicide less harshly than intended killing.⁵⁶ At common law, premeditated intentional killings are treated as first degree murder and punished more harshly than premeditated ones done with merely the awareness of an unjustifiable risk (depraved

subjectivist theory of punishment. Alexander and Ferzan would punish equally actors who (1) desired to cause a harm and acted in a manner creating a risk and (2) actors who merely acted with the knowledge of recklessly creating the risk. ALEXANDER & FERZAN, *supra* note 5, at 38–39.

56. Compare MODEL PENAL CODE § 210.2 (1962), *with id.* § 210.3. The Model Penal Code allows a small category of reckless killings to qualify as murder when such recklessness can fairly “be assimilated to purpose or knowledge.” *Id.* § 210.2 cmt. 4 (Official Draft and Revised Comments 1980).

heart killings), which are treated as second degree murders.⁵⁷ But between $R \times H$ of Case 4 and H of Case 1, there seems to be no principled rule for determining the appropriate sanction for Case 3.

A similar argument to the same end can be made by first considering Case 3. As argued above, it seems reasonable that if a person, as in Case 1, intentionally causes a harm, his culpability should be equivalent to a person who knowingly causes a harm. In both cases, their sanction should be equivalent to H . A person who knowingly causes harm is roughly a person who believes there is a 100% chance harm will occur.⁵⁸ Robert Nozick famously claimed that desert equals degree of responsibility times harm ($D = r \times H$), where responsibility could range from 0 to 1.⁵⁹ The greater the risk created of H , the greater the responsibility for H is realized. Assume, as might be true in Case 3, an actor believes there is a 50% risk that the harm will occur. In that case, the actor's culpability seems half that of the person who believes the harm has a 100% chance of occurring. As a general matter, a sanction equivalent to $R \times H$ in Case 3 seems just.⁶⁰

But what then should the sanction be for Case 4? There's the rub. According to the harm-based retributivism proponent, the sanction should be less than in Case 3 because of the absence of harm. But there seems no principled way of determining a sanction less than $R \times H$. In particular, the differential applied between Case 1 and Case 2—having the sanction depend on harm-producing conduct rather than risk-producing conduct—is not available to be applied to Cases 3 and 4 because the sanction in Case 3 already turns on the risk of harm; it is not possible to *shift* to risk of harm as the basis of punishment in Case 4. It might be suggested that the sanction in Case 4 be $(R \times R)H$ on the ground that the sanction for Case 3 should be $R \times H$, and discounting the

57. See LAFAYE, *supra* note 36, §§ 14.4, 14.7.

58. "Knowledge" is often defined in the criminal law as "practical[] certain[ty]." MODEL PENAL CODE § 2.02(2)(b)(ii) (1962). Because certainty is fairly construed as complete or 100% conviction in the truth of a proposition, a person who is practically certain believes there is roughly, but not quite, a 100% chance that the proposition is true.

59. NOZICK, *supra* note 3, at 60.

60. Acting with the belief that one's conduct carries a risk of harm thus may have moral significance in two ways. First, where the conduct fails to produce harm, the belief provides a moral basis for characterizing the conduct as a wrongdoing. Second, where the conduct is not intended to produce harm, but does, the belief provides a basis for holding the person responsible for the wrongdoing, now characterized as harm-causing conduct.

sanction by R , as done in Case 2, correctly reflects the distinction between cases of harm and no harm. Such a position, besides implying the surprising identity of punishments in Cases 2 and 3, has questionable appeal. $R \times H$ made sense for Case 2 because it reflected the risk of harm that was intentionally imposed on B . $R \times H$ made sense in Case 3 because it reflected a lessened responsibility for an actual harm. The natural interpretation of a penalty of $R(R \times H)$ would be for conduct that the actor believes has an R likelihood of creating a risk R . Thus there would be diminished responsibility for creating risk R . But A in Case 4 knows the risk she is imposing, and so should take full responsibility for it. As argued above, $R \times H$ seems the appropriate punishment for Case 4.

III. AN ALTERNATIVE TO HARM-BASED RETRIBUTIVISM

To recap: the difficulty with harm-based retributivism is that although it seems most naturally to suggest a sanction equivalent to H in cases of intentional harm (Case 1), and seems to suggest a principled extension (the harm/risk theory) for the punishment of attempts (Case 2), it does not seem capable of being elaborated in a manner in which the sanctions for unintended realized risks and unintended unrealized risks (Case 3 and Case 4) are both nonarbitrary and respect the difference between harm and no-harm cases in a manner paralleling the difference in sanctions in cases where harm is intended (Cases 1 and 2). The theory advanced below, modal retributivism, avoids this difficulty while remaining faithful to the intuition underlying harm-based retributivism that resultant harm matters to desert.

A. *Positive and Negative Retributivism*

Thus far I have focused on one distinction between types of retributivist theories: harm-based versus intent-based retributivism. Another distinction may be drawn: positive retributivism and negative retributivism. The positive/negative retributivism distinction and the harm-based/intent-based retributivism distinctions are independent. A position on one does not imply a position on the other.

According to positive retributivism, those who engage in culpable wrongdoing *should* be subjected to a sanction, the magni-

tude of which is determined by a theory of desert.⁶¹ By this positive retributivists mean that culpable wrongdoing by an actor is a ground, and provides a reason, for harsh treatment of the actor. While not necessarily mandatory,⁶² punishment, where deserved, is precatory. Different positive retributivists may have different views on the strength and nature of the ground or reason.⁶³

In contrast, according to negative retributivism, those who engage in culpable wrongdoing *may* be subjected to sanctions up to a level determined by a theory of desert and no more.⁶⁴ Conditions for desert thus are regarded as necessary, but never sufficient, grounds for punishment. The conditions establish an upper limit of punishment. Often negative retributivism is supplemented by a consequentialist or other theory of punishment that determines when a person, who may be punished pursuant to negative retributivism, should be punished.⁶⁵ (For purposes of illustration, I shall generally assume a consequentialist supplement to negative retributivism along simple utilitarian lines.) Such an approach combines deontological and consequentialist approaches to punishment and is sometimes called a "mixed" (or "hybrid") theory of punishment.⁶⁶ A proponent of such a mixed theory might hold that a murderer may be executed by virtue of his culpably causing a death, and, given this, should be executed on the ground that doing so would maximize social utility.⁶⁷ Neg-

61. See Christopher, *supra* note 7, at 865–66.

62. A stronger position—one that might be called superpositive retributivism—would assert that in cases of culpable wrongdoing punishment is mandatory, and should be imposed no matter what the cost to society and regardless of how such punishment might conflict with other moral norms. Such a view might require, for example, the sacrifice of innocent lives or significant losses in social welfare necessary to ensure that those who deserve to be punished are punished. It is doubtful that any contemporary theorist holds such an extreme view. See *infra* Part IV.

63. Moore, for example, believes that moral culpability for breaking society's law establishes a (presumably defeasible) moral duty of society to punish. Moore, *supra* note 1, at 182.

64. See Crocker, *supra* note 21, at 1060–61 (defending this position); Alan H. Goldman, *The Paradox of Punishment*, 9 PHIL. & PUB. AFF. 42, 44 (1979). See generally H.L.A. HART, *Prolegomenon to the Principles of Punishment*, in PUNISHMENT AND RESPONSIBILITY 1–27 (1968) [hereinafter HART, *Prolegomenon*] (advocating such a position); H.L.A. HART, *Punishment and the Elimination of Responsibility*, in PUNISHMENT AND RESPONSIBILITY, *supra*, at 159–85 (same).

65. See Paul H. Robinson, *Hybrid Principles for the Distribution of Criminal Sanctions*, 82 NW. U. L. REV. 19, 33–34 (1988).

66. See *id.* at 36; Andrew von Hirsch, *Hybrid Principles in Allocating Sanctions: A Response to Professor Robinson*, 82 NW. U. L. REV. 64, 64 (1988).

67. See Cederblom & Spohn, *supra* note 6, at 921 (advocating such a theory); cf. HART, *Prolegomenon*, *supra* note 64, at 3 (arguing that "a plurality of different values

ative retributivism is thus a theory of moral forfeiture according to which culpable wrongdoing results in a limited forfeiture of the right not to be subject to sanctions for other reasons.

The discussion to this point has employed language implying positive retributivism. The difficulty harm-based retributivists have in specifying sanctions for inchoate offenses, however, apply equally to positive and negative harm-based retributivists. Just as positive harm-based retributivists must struggle to specify the magnitude of sanctions that should be applied to those who have committed inchoate offenses, so must negative harm-based retributivists struggle to specify the magnitude of sanctions that may be applied to those who have committed inchoate offenses.

B. *Modal Retributivism*

In a nutshell, my proposal is that when there is harm, positive retributivism should apply, but where there is no harm, negative retributivism should apply. Because the mode of punishment switches from “should” to “may” based on whether there was harm, the theory of punishment may be described as “modal retributivism.”⁶⁸ It is summarized in the table below:

and aims . . . are relevant at different points in any morally acceptable account of punishment”).

68. Theories of punishment that assert that an actor’s culpability licenses sanctions and that consequences of punishment justify sanction (but only to the point they are licensed) are sometimes called “mixed theories” of punishment. Because modal retributivism asserts that culpability merely licenses sanction in some cases, but not others, modal retributivism could be described as a partial-mixed theory of punishment.

Modal Retributivism

	Harm <i>H</i>	No Harm
Harm <i>H</i> intended; <i>R</i> risk of <i>H</i>	Case 1 Example: Murder Sanction: <i>H</i> should be imposed	Case 2 Example: Attempted murder Sanction: <i>H</i> may be imposed
Harm <i>H</i> unintended; <i>R</i> risk of <i>H</i>	Case 3 Example: Reckless homicide Sanction: <i>R</i> x <i>H</i> should be imposed	Case 4 Example: Reckless endangerment Sanction: <i>R</i> x <i>H</i> may be imposed

The initial appeal of modal retributivism lies in the following features: It presents a complete theory of punishment in the sense that it prescribes the severity of sanction to be applied for consummate and inchoate crimes alike. It distinguishes between cases where harm is caused and where it is not. It does so in a consistent manner, giving the presence or absence of harm the same effect in the Case 1 and Case 2 pair as in the Case 3 and Case 4 pair. The distinction it draws between precatory and permissive sanctions is not facially arbitrary in the manner that sanctioning attempts at 7/9 the level of a consummate offense would be.

Moving beyond the initial appeal, though, the question is whether the precatory/permissive distinction modal retributivism draws between cases of harm and no harm is the right distinction. As noted earlier, matters of morality are not susceptible to demonstration by experimentation or proof. Often a valid argument for a moral position is simply a claim that it is plausible and that there are no plausible alternatives. I have tried to show that the alternatives to modal retributivism are not plausible. Are they?

Clearly the components of modal retributivism, considered in isolation, are plausible. Positive retributivists would agree that positive retributivism is appropriate in cases where harm is caused. Negative retributivists would agree that negative retributivism is appropriate in cases where harm is not caused. And they would both agree that in cases of harm, the actor, at a minimum, may be sanctioned to a point equivalent to the harm imposed. Here the negative retributivist might explain that punishment is permissible in cases where the intended harm occurs, as in cases where the intended harm fails to occur, because in intending harm, the actor manifests his belief that it is acceptable to use another as a means to an end—at least to a degree. Thus, the actor is not in a position to object to being used as a means to an end by society to that degree. The actor has voluntarily forfeited her right not to have harsh treatment imposed on her. In such cases, punishment is permissible and, where it advances society's ends—the deterrence of crime, for example—it is appropriate.⁶⁹

The critical point is whether there is a stronger case for positive retributivism in cases of intended harm than in cases where harm was intended but did not occur. Positive retributivists believe that even in the absence of consequential benefits to society, sanctions should be imposed.⁷⁰ At this point, different positive retributivists may tell different stories to support or explain this very basic moral intuition. For example, it may be thought that punishment should be imposed in order to restore in some sense the moral balance of the universe. Where a harm has been done to an innocent victim by a culpable actor, imposing an equally severe harm on the actor culpable for the harm restores the balance.⁷¹ Or, to paraphrase the Golden Rule, as you treat others, so they will treat you. Colloquially, what goes around should come around. A person should get a taste of his own medicine. Such a story appears much less cogent in cases where no harm has been caused. Consider Case A, where *A* intends to brutally assault a child and does so, and Case B, where *B* acts on his intent to bru-

69. See Crocker, *supra* note 21, at 1072–79 (arguing for negative retributivism as a consequence of the more general principal of reciprocity); Goldman, *supra* note 64, at 42–47 (explaining forfeiture theory).

70. Eugene E. Dais, Commentary, *Positive Retributivism and Despicable Justice*, in *RETRIBUTIVISM AND ITS CRITICS* 107, 107 (Wesley Cragg ed., 1992).

71. Christopher, *supra* note 7, at 860 (“It is morally fitting that an offender should suffer in proportion to her desert or culpable wrongdoing.”).

tally assault a child and the child escapes unharmed. The difference in our reaction is not that we feel as strongly in the two cases that *B* should be punished, but punished less; but, I suggest, that we feel less strongly that *B* should be punished. The retributive impulse is dampened. The idea of punishment for its own sake, the defining feature of any nonconsequentialist theory, has a much weaker appeal.

Useful comparisons may be drawn between retributive justice and other forms of justice. Where, and only where, an actor actually harms another does the victim deserve to be compensated.⁷² Likewise, where and only where an actor has actually created an item through his own labor does the actor deserve possession and control of the item.⁷³ Likewise, where and only where an actor has actually profited through his wrongful act does the actor deserve to forfeit his unjust enrichment.⁷⁴ With respect to both corrective justice, property entitlements, and unjust enrichment, certain consequences of conduct are necessary conditions to trigger precatory claims of desert. Punishment for inchoate offenses, where conduct need not have morally relevant consequences, seems an odd candidate for precatory punishment. In my view, it is the weak sister in the family of conditions to be treated according to justice. In order to be supplemented it needs some tie to effects in the real world. That tie, however, is not the effects of past conduct, but to the future socially beneficial effects of punishment. Without these beneficial effects, punishment is not precatory. Rather, punishment is only permitted on the condition that such beneficial effects will occur. This is exactly the theory of modal retributivism.

C. *Modal Retributivism and Plato*

To my knowledge, no theorist of punishment has taken a position which strictly instantiates modal retributivism. However, modal retributivism is not so far from existing theories as to make it outlandish. A theory of punishment along the lines of

72. See VIVIENNE HARWOOD, *PRINCIPLES OF TORT LAW* 1–3 (4th ed. 2000).

73. JOHN LOCKE, *TWO TREATISES OF GOVERNMENT* 288–89 (Peter Laslett ed., Cambridge Univ. Press 1988) (1690).

74. See Andrew Kull, *Restitution's Outlaws*, 78 CHI.-KENT L. REV. 17, 18 (2003) (explaining that restitution is intended to strip the actor of any profit he has gained as a result of his wrongful conduct).

modal retributivism can be found, for example, in Plato's discussion of criminal punishment, as elaborated by Peter Westen.⁷⁵

Plato's views on punishment are not deeply theorized by Plato himself. Rather they must be inferred from the penalties he recommends and his brief attendant comments. First, the penalties he recommends are in line with those recommended by modal retributivism. Plato believes those who kill intentionally and unlawfully should be put to death.⁷⁶ This is consistent with the equality (*lex talionis*) component of modal retributivism. Regarding the question whether those who try to kill, and do not, should be put to death, Plato's answer is: they may, it depends.⁷⁷ According to Plato, whether they should be turns on the relationship of the would-be killer and the intended victim.⁷⁸ While in the general case execution is inappropriate, in cases of patricide and the killing of a master by a slave, execution is appropriate.⁷⁹ Thus, Plato's penalty recommendations may be characterized as precatory when there is resultant harm, and permissive (depending on the status of the parties) in cases of attempt. Plato's theory thus tracks modal retributivism.

Second, the rationale behind these recommendations seems a form of modal retributivism. Why, Westen asks, does Plato make these distinctions, first between those who cause harm and those who do not, and second among those who do not?⁸⁰ According to Plato, the difference in treatment between killers and attempted killers is due to the fact that the sparing of the life of the latter is a "thank-offering" to the deity responsible for the fact no death occurred.⁸¹ Westen interprets Plato as saying that where an attempt to harm is unsuccessful, society should feel gratefulness and only a muted indignation, and that in this light, a reduced punishment is appropriate.⁸² Stated conversely, it is the indigna-

75. Peter Westen, *Why Criminal Harms Matter: Plato's Abiding Insight in the Laws*, 1 CRIM. L. & PHIL. 307, 315–18 (2007).

76. Plato, *Laws*, in THE COLLECTED DIALOGUES OF PLATO, INCLUDING THE LETTERS, 1225, 1432 (Edith Hamilton & Huntington Cairns eds., A.E. Taylor trans., Bollingen Found. 1961).

77. *Id.* at 1435–36.

78. *Id.*

79. *Id.*

80. Westen, *supra* note 75, at 315–16.

81. Plato, *Laws*, in 5 THE COLLECTED DIALOGUES OF PLATO 261–62 (B. Jowett ed., Macmillan and Co. 1892).

82. *See* Westen, *supra* note 75, at 318.

tion that society rightly feels when death occurs, along with the offender's personal culpability, that properly motivates society's extracting full punishment. Westen is explicit that in Plato's view, the fact of harm is not related to the desert of the offender,⁸³ as it would be pursuant to strict modal retributivism. Nevertheless, the critical fact is that harm justifies punishment for non-consequentialist reasons as in modal retributivism.

In contrast, where there is no death, the status relationship of the parties matters. In the absence of an explanation by Plato, Westen theorizes that where the attempt to kill fails, the actor should be executed either because (1) the act is so heinous that even when the mitigation based on gratitude for survival is taken into account, execution is appropriate; or (2) patricide and slave/master murder are so socially destructive that society's interest in deterrence overrides "the value of the supererogatory act of remitting the death penalty in the event of failure."⁸⁴ It is this second explanation that interests me. If it is correct, then the significance of death, or harm more generally, is that when it is culpably caused, no further justification is needed; in contrast, where harm is absent, society has no intrinsic reason to act, and consequentialist considerations come into play to determine the level of punishment. This is the defining feature of modal retributivism: where there is harm, nonconsequential considerations justify full punishment; where there is no harm, consequential and deontological ones jointly determine the level of punishment.

D. *Modal Retributivism Elaborated*

The previous sections discussed the operation of modal retributivism in the context of three types of *mentes reae*: purpose, knowledge, and risk awareness. The underlying thesis was that it made sense and was possible to produce a deontological theory that, with the appropriate empirical inputs, such as the magnitude of harm and level of subjective risk, prescribed precise punishments across a range of offenses. This section shows how this thesis might be extended to offenses based on heat of passion and negligence.

83. *Id.*

84. *Id.*

1. Crimes of Passion

Voluntary manslaughter is a crime of passion.⁸⁵ Loosely speaking, a person commits voluntary manslaughter when she kills another in a sudden state of emotional turmoil in response to adequate provocation.⁸⁶ The penalty for voluntary manslaughter is less than that for murder, but generally equal to or greater than the penalty for involuntary manslaughter.⁸⁷ What should the penalty be for voluntary manslaughter, consistent with the goal of this article to develop a theory of nonarbitrary sanction recommendations? In particular, how should the sanction level compare with the penalty for murder?

Before answering this question, it should be noted that voluntary manslaughter is unique in providing a mitigated sanction for an intentional wrongful act taken in the heat of passion.⁸⁸ If an attack results in severe permanent injury, but not death of a victim, the fact that it was committed in the heat of passion is irrelevant to the determination of the sanction range the offender may be exposed to.⁸⁹ There seems no principled reason why heat of passion should be relevant in one case, but not the other. The criminal law thus seems internally conflicted regarding whether acting in the heat of passion should be a mitigating factor, at least at the level of offense definition. Almost all commentators, however, endorse mitigation for homicide,⁹⁰ and so I shall offer an approach to determining the sanction for manslaughter. The approach, in theory, could be extended to other offenses committed in the heat of passion.

There is ongoing controversy whether the penalty discount for manslaughter relative to murder is better understood as a partial justification based on the wrongfulness of the provoking event or a partial excuse based on justified emotional turmoil.⁹¹ Justifica-

85. LAFAVE, *supra* note 36, § 15.2.

86. *Id.*

87. *See id.* § 15.1

88. *See id.* § 15.2(h).

89. *See id.*

90. An exception is Cynthia Lee, who advocates the doctrine's repeal. *See* CYNTHIA LEE, MURDER AND THE REASONABLE MAN (2003).

91. *See* Victoria Nourse, *Passion's Progress: Modern Law Reform and the Provocation Defense*, 106 YALE L.J. 1331, 1339–40 (1997) (discussing the Model Penal Code's emphasis on provocation as a partial excuse based on the defendant's emotional disturbance, rather than a partial justification based on the provocation by the victim); *see also* Kyron Hui-

tion defenses are appropriate in cases where the wrongdoing is either annulled or offset by the existence of justifying circumstances. If H is the harm done in a killing, what is the harm done in a case of voluntary manslaughter when some of the wrongdoing is either annulled or offset by the partially justifying circumstances of provocation? $(2/3)H$? $(3/4)H$? Excuse defenses are appropriate in cases where responsibility for a wrongdoing is diminished. According to the equality principle, in the typical case an actor is maximally responsible where she intended the harm or knew the harm would be caused. Where an actor merely risks causing a harm, her responsibility for the harm is less. I have argued that in such cases, the sanction should reflect this lesser responsibility by being $R \times H$, where R is less than 1. What diminution in responsibility does acting in the heat of passion entail? $(2/3)H$? $(3/4)H$? Neither a justification approach nor an excuse approach appears to offer a framework for deciding the amount of the discount for killing in the heat of passion.

Without attempting to characterize my approach as either justification based or excuse based, let me offer the following thoughts on voluntary manslaughter. When one kills another in the heat of passion, one knows that one is killing another person. Neither intense fear, anger, nor jealousy interferes appreciably with an actor's cognitive faculties. Instead, intense fear, anger, jealousy, and the like interfere with an actor's emotional faculties. Though she knows that the provoker she is killing is human, the killer acting in the heat of passion does not feel the significance of this. The revulsion, incipient sorrow, guilt, and moral gravity associated with the prospect of killing another is obscured by the overwhelming feelings of fear, anger, or jealousy the actor experiences. Potential empathy for the victim is suppressed. Killing when provoked does not *feel* like the killing of another. On the emotional level, it feels as if one is committing a much less harmful act. Of course, in order for the actor to get the benefit of the manslaughter discount, this distorted feeling must be reasonable under the circumstances. The actor cannot be blamed for her failure to feel as she should. In such circumstances, the sanction that should be imposed on the actor should be equal in magnitude to the harm she (reasonably) feels she is causing, or, in

gens, *Homicide in Aretaic Terms*, 6 BUFF. CRIM. L. REV. 97, 132-45 (2002) (advocating an abolishment of "provocation" excuse because it is "pernicious and unnecessary").

other words, equal in magnitude to the harm of the type that a reasonable person in her position would feel like she is causing. This will vary based on the intensity of the passion that the actor appropriately felt. Thus, the theoretically ideal sanction for manslaughter would be a sliding scale based on the evaluation of a counterfactual proposition about the degree of wrongfulness a reasonable person would feel.⁹²

The question then arises: to what extent is the reasonable person prone to having her normal moral perceptions of the value of others obscured by emotions like fear and anger? Anger and fear sometimes serve a valuable purpose. When we are attacked, for example, anger can free us from the squeamishness and second thoughts that might otherwise interfere with our own self-defense. Similarly, fear promotes justified risk avoidance by acting as a counterweight to greed, gratification, and the tendency to discount future harms. Furthermore, living a life that fosters emotional control—for example, the life of a Buddhist monk or a strict military career—at least risks an unhealthy detachment from the richness of life or a limited emotional capacity for empathy and expression. It has been theorized that reasonable persons take risks to the extent the risks maximize social welfare.⁹³ Likewise, it may be argued reasonable persons develop the capacity for emotions like fear and anger to the extent those emotions maximize social well-being, however conceived. Although this amount cannot practically be determined with any higher degree of accuracy than the reasonable level of risk taking, it is theoretically specifiable given sufficient empirical information and, once specified, implies—as argued above—a principled punishment level for voluntary manslaughter.

2. Crimes of Negligence

Finally, it may be asked where offenses of negligence, such as negligent homicide or battery, fit into modal retributivism.⁹⁴

92. A similar approach of understanding crimes of passion as involving undiminished culpability for lesser harms, rather than diminished culpability for full harms, is taken by Cederblom and Spohn. Cederblom & Spohn, *supra* note 6, at 925–26.

93. See RICHARD A. POSNER, *ECONOMIC ANALYSIS OF LAW* §§ 6.1–6.2 (7th ed. 2007).

94. Criminal offenses of negligence usually require a higher level of negligence than simple or civil negligence, sometimes referred to as “criminal” or “gross” negligence. Regardless of degree, the defining feature of criminal negligence offenses is that they do not require awareness of the unacceptable risk created, as do offenses of recklessness. See,

When it comes to the criminal law, this question is not directly applicable. The central thesis of modal retributivism is that cases where harm is caused should be punished and cases where there is no harm may be punished. Currently, there are no offenses of negligent conduct that do not result in harm.⁹⁵ Thus, strictly speaking, the application of modal retributivism to offenses of negligence does not arise.

Nevertheless it may be asked, according to the general approach this article has followed to issues of punishment: what is the appropriate punishment for acts of negligence that cause harm? This article has argued that the sanction for recklessly caused harms should be $R \times H$, where R is the subjective risk of harm and H is the magnitude of harm. It is universally agreed among commentators that negligence is a lesser form of culpability than recklessness, and that therefore acts of negligence that result in a given harm deserve less severe sanctions than reckless acts causing equivalent harms.⁹⁶ So if negligence deserves to be punished, some principled amount of punishment between 0 and $R \times H$ needs to be found.

With respect to manslaughter, there was widespread agreement that killings due to the reasonably inflamed passions of the actor should be punished, though at a lesser level than cold-blooded or inadequately provoked killings. In contrast, there is substantial debate whether those who negligently cause harm should be subjected to criminal sanctions at all.⁹⁷ Some take the position that there should be no criminal offenses based on negli-

e.g., MODEL PENAL CODE § 2.02(2)(c)-(d) (1962).

95. It is an interesting question why a person who engages in negligent conduct, but luckily fails to cause harm, should escape criminal liability while a person who negligently causes harm, such as death, may be subjected to significant criminal penalties. This is true even in the Model Penal Code, which generally established the same level of punishment for consummate offenses and attempts. *See id.* § 5.05(1).

96. *See, e.g.*, LAFAYE, *supra* note 36, § 5.4. In contrast, the common law provides some support for the proposition that negligence should be punished equally with intent and recklessness. Under the common law, various defenses, such as self-defense and necessity, are not available if the actor negligently believed the conditions for the defense existed. For example, an actor could not claim self-defense if she negligently believed the victim was threatening her. *See* DRESSLER, *supra* note 4, § 18.01[E], at 225. The actor would be liable as much if he had been reckless with respect to, or had known of, the absence of these conditions. The Model Penal Code rejects this approach and treats negligence regarding defense conditions like negligence regarding offense conditions—lesser liability compared to recklessness and knowledge results. *See* MODEL PENAL CODE §§ 3.02(2), 3.09(2).

97. *See, e.g.*, DRESSLER, *supra* note 4, § 10.04[D][2][c].

gence.⁹⁸ They argue that the moral culpability associated with negligence is not great enough to warrant criminal sanctions.⁹⁹ As a normative matter, the argument is that there is an important moral distinction between being lazy, careless, unthinking, or unintelligent, and being evil.¹⁰⁰ As a descriptive matter, they might point to the fact that there are no criminal offenses of negligent harm to property and that frequently there are no offenses based on the negligent causing of nonfatal injuries. Given that moral intuition supports a lesser sanction for negligently caused harm, and that there is reasonable disagreement whether negligently caused harm should be punished at some level less than $R \times H$ or not at all, I advocate the latter position. The reason is that there seems to be no principled basis for determining the amount of a sanction that lies between $R \times H$ and 0. There appears to be neither an attractive fixed point, such as $R \times (H/2)$, nor a relevant factor that might establish a sliding scale for a criminal sanction. Until a theory is advanced,¹⁰¹ it seems to make more sense to adopt the position that tort liability for negligently caused harms satisfies the demands of justice. Here, for once, zero punishment seems the appropriate principled criminal sanction.

IV. IMPLEMENTING MODAL RETRIBUTIVISM

Assuming modal retributivism and the other sanction schemes recommended herein are found appealing, various changes in existing law would have to be made to implement them. Four such changes are described below.

First, the most obvious change would be to equalize the maximum penalties for consummate offenses and attempts. Current

98. *Id.*

99. See, e.g., ALEXANDER & FERZAN, *supra* note 5, at 71 (“[A]cts of clumsiness and stupidity hardly seem to be the sort of things for which we wish to hold people criminally liable . . .”); Jerome Hall, *Negligent Behavior Should Be Excluded From Penal Liability*, 63 COLUM. L. REV. 632 (1963).

100. ALEXANDER & FERZAN, *supra* note 5, at 70–71.

101. Perhaps a theory may be advanced along the following lines: Sometimes persons unreasonably fail to perceive that their conduct creates an unjustifiable risk of harm when there is strong evidence for finding an unjustifiable risk; others do so where there is weaker but still sufficient evidence. The former are arguably more negligent and more culpable. Perhaps there is a means of calibrating the unreasonableness of failing to perceive the risk by assessing the strength of the evidence for the risk and, on this basis, determining, along with the magnitude of the risk and the amount of harm, the appropriate level of sanction.

law establishes lesser sanctions for attempts. Under modal retributivism, a sanction of H or $R \times H$ (for offenses of recklessness) may be imposed even in the absence of harm, depending on whether there is an independent consequentialist justification for the sanction. Assuming that appropriate sanction levels are already in place for consummate offenses, the maximum penalty levels for attempts would have to be raised to those levels.

Implementing such a change directly might not be possible for attempted murder. Although it has never squarely considered the question, it is likely that the Supreme Court would hold that executing an actor for attempted murder would violate the constitutional prohibition of cruel and unusual punishment.¹⁰² Thus, under current law, an actor who luckily fails to kill his intended victim may not be executed, and another actor who does kill his intended victim would be executed, though both (1) demonstrate through their conduct an utter disregard for the sanctity of life and (2) present a relatively high probability of killing an innocent person in the future unless executed. Given this constitutional constraint, modal retributivism might recommend as a possible maximum sentence for attempted murder a sanction like life imprisonment in solitary confinement. Such a sanction is arguably as harsh as the death penalty.¹⁰³

Second, minimum penalties for attempts should be lowered to a de minimis amount to allow for cases where there is no independent consequentialist reason to impose a sanction. According to modal retributivism, where an actor has recklessly engaged in conduct or has tried to cause harm, yet no harm has ensued, punishment is merely contingently licensed. In some cases, for example where the actor demonstrates no likelihood of recidivism

102. The Supreme Court's most recent pronouncement on the scope of the death penalty can be found in *Kennedy v. Louisiana*, 554 U.S. ___, 128 S. Ct. 2641 (2008). In *Kennedy*, the Court held that the Constitution forbids execution for child rape. *Id.* at ___, 128 S. Ct. at 2646. The Court stated therein, "[O]ur holding is that a death sentence for one who raped but did not kill a child, and who did not intend to assist another in killing the child, is unconstitutional under the Eighth and Fourteenth Amendments." *Id.* at ___, 128 S. Ct. 2650-51. This statement seems ambiguous. It might be read to recognize the constitutional possibility of capital punishment in either cases of (a) child rape where there was an intent to assist in a murder that did not occur, or (b) child rape-plus-murder-by-another where there was an intent to assist in the murder. Under the first interpretation at least sometimes capital punishment for attempted murder would be constitutional; under the second interpretation, not. Either interpretation, of course, is dictum at best.

103. See David McCord, *Imagining a Retributivist Alternative to Capital Punishment*, 50 FLA. L. REV. 1, 104-09 (1998).

and punishment will not deter, there will be no independent reason to impose a sanction. Likewise, there may be cases where relatively weak gains in deterrence and other benefits are offset by significant social costs of incarceration. Here no sanction is appropriate. Of course, in many cases of criminal attempts, there will be some harm, such as anxiety and apprehension, caused to the intended victim or to the community. In those cases, some punishment would be called for, at least to the extent the harm was foreseen. Some *de minimis* punishment would likely be warranted.

By the same token, many offenses that have significant inchoate aspects should have minimum penalties lowered to a *de minimis* level. By “offenses that have significant inchoate aspects,” I refer to offenses with maximum penalty levels that cannot be justified in harm-based retributivism terms by appealing to harms required by the offense’s definition. An example of such an offense is larceny. In common law-based jurisdictions, larceny is frequently defined as “the trespassory taking and carrying away of the personal property of another with the intent to permanently deprive the possessor of the property.”¹⁰⁴ If the property is worth over \$500, the minimum penalty for theft is a year incarceration.¹⁰⁵ An actor can commit a larceny by causing only minor interference with the use of an item.¹⁰⁶ Removing an item only for a minute or a meter satisfies the requirements of larceny. Causing such a fleeting harm could not in itself justify imposing a sanction of a year in jail. Rather than an offense of harm causing, larceny is better thought of as a specific type of criminal attempt—attempting to deprive a person permanently of property.¹⁰⁷ Thus, it should have a penalty range similar to an attempt: from a *de minimis* penalty to a penalty equivalent to the harm associated with cases where a permanent deprivation was in fact accomplished. Perjury is another crime defined in the inchoate mode. The harm that it seeks to prevent is erroneous verdicts, but perjury may be committed even if the perjury is not believed by the decisionmaker, or even if it does not contribute to an erroneous verdict.¹⁰⁸ A similar analysis would apply to many other of-

104. DRESSLER, *supra* note 4, § 32.02[A].

105. See MODEL PENAL CODE § 223.1(2)(a) (1962).

106. See, e.g., DRESSLER, *supra* note 4, § 32.05.

107. See DUFF, *supra* note 4, at 128–29 (discussing implicit inchoate offenses).

108. Cf. MODEL PENAL CODE § 241.1(2) (defining the material false statement required

fenses, including burglary and most possessory offenses. Other offenses with significant inchoate aspects, such as rape,¹⁰⁹ should also have minimum penalties lowered.

Third, the negligent causing of death should be treated like the negligent causing of injury and the negligent causing of property damage.¹¹⁰ There should be no criminal penalty; rather, liability should be civil only.

Fourth, nonbinding guidelines should be legislatively promulgated. Thus far this discussion has assumed the standard offense definitions and has advocated a range of sanctions wide enough to accommodate those cases satisfying the offense definition. In theory, the offenses themselves could be defined in a manner to more closely track the sanction recommendations of modal retributivism. For example, a hypothetical offense of "first degree attempted arson" might be defined as "attempted arson by an actor whose lengthy incarceration would produce a net benefit for society" and carry a penalty equal to that of arson. The difficulty with such an approach, of course, is that jurors are not well positioned to determine whether consequentialist-related conditions such as being "an actor whose lengthy incarceration would benefit society" are present. Jurors do not have the training or experience to make competent cost-benefit determinations.

The alternative that establishes a wide sentencing range of attempt sanctions for judges to choose from, however, cannot be legislatively supplemented to ensure that judges will act consistently with modal retributivism. Recent Supreme Court case law bars legislatures from establishing sentencing guidelines that set conditions that a judge must find in order to impose an enlarged

for perjury as one that "could have affected the course or outcome of the proceeding" (emphasis added)).

109. The physical injuries caused by rape are often less severe than in cases of aggravated assault. The very high penalties for rape are thus best justified in retributivist terms by the severe emotional and psychological harm that typically result from rape. These emotional and psychological harms, however, are not a required element of rape. Under modal retributivism, in the atypical case where such harm, even though culpably risked, does not occur, punishment should not exceed an amount equal to the actual physical harm.

110. The Model Penal Code creates no liability for negligent physical injury (unless done with a deadly weapon) or property damage. See MODEL PENAL CODE § 211.1(1)(b) (1962). Under the common law, there generally was no criminal liability for negligent property damage, and no clearly established criminal liability for negligently caused physical injury. LAFAVE, *supra* note 36, §§ 16.2(c)(2).

sanction.¹¹¹ A judge, for example, cannot be instructed by the legislature to only impose the maximum sentence where there is a consequentialist justification to do so; the judge must be given discretion to impose it on the ground (contrary to modal retributivism) that the actor's purpose was evil. For this reason, the negative retributivist side of modal retributivism, which asserts that evil purpose is not a reason to impose punishment, is difficult to implement directly through legislation. Nonbinding guidelines, however, are constitutional, and may be used to encourage judges to set sentences along modal retributivist lines.¹¹²

V. QUALIFICATIONS AND CONSIDERATIONS

A few closing qualifications and considerations are in order.

First, this article has not attempted to defend from the ground up either retributivism or the intuition that resultant harm matters for punishment. These sentiments have wide appeal and have been examined by others. Nor has modal retributivism been advanced as the exclusive determinant of punishment. For example, remorse of the offender, forgiveness of the victim, mercy of the authority establishing the norm transgressed, or conflicting principles of justice, may properly bear upon punishment.¹¹³ The last twenty years have seen a flourishing of intriguing non-retributive punishment theories, such as expressivism, atonement, and restorative, Aristotelean, communicative, and victim vindication theories.¹¹⁴ These theories might be incorporated into

111. See *United States v. Booker*, 543 U.S. 220, 244 (2005) (holding mandatory application of federal sentencing guidelines unconstitutional); *Blakely v. Washington*, 542 U.S. 296, 301–05 (2004) (holding that the jury's findings, and not additional findings by the judge, constitute the allowable statutory maximum sentence a judge may impose under a state sentencing guidelines scheme); *Apprendi v. New Jersey*, 530 U.S. 466, 490–92 (2000) (holding that "any fact that increases the penalty for a crime beyond the prescribed statutory maximum must be submitted to a jury, and proved beyond a reasonable doubt," excepting the fact of a prior conviction).

112. See sources cited *supra* note 111; see also John E. Pfaff, *The Continued Vitality of Structured Sentencing Following Blakely: The Effectiveness of Voluntary Guidelines*, 54 UCLA L. REV. 235 (2006) (presenting empirical evidence that Federal Sentencing Guidelines, even though nonbinding, significantly influence sentencing decisions of federal judges).

113. The author has argued principles of distributive justice are relevant in determining the punishment due to entrapped offenders. See Anthony M. Dillof, *Unraveling Unlawful Entrapment*, 94 J. CRIM. L. & CRIMINOLOGY 827, 827–31, 871–75 (2004).

114. See generally R.A. Duff, *Penal Communications: Recent Work in the Philosophy of Punishment*, 20 CRIME & JUST. 1 (1996) (discussing the communicative theory of punishment); Stephen P. Garvey, *Punishment as Atonement*, 46 UCLA L. REV. 1801, 1813 (1999)

modal retributivism by providing the positive reason for punishment in cases where modal retributivism merely licenses punishment. Alternatively, the punishment recommendation of these theories might have to be balanced against those of modal retributivism. Consequentialist considerations, for example, are likely always to be included in the balance to some degree. I know of no contemporary theorist who takes the position that a person must be punished in accordance with his desert no matter what the cost.¹¹⁵ Most, if pressed, I believe would agree that even a person who culpably killed another without excuse or justification need not be punished if the life of an innocent person could thereby be saved without other cost to society. Likewise most, if pressed, I believe would agree that even an innocent person might be subjected to criminal sanctions if, for example, the lives of one thousand people could directly and immediately be saved by it.¹¹⁶ Accordingly, this article should be viewed as resting on the relatively modest premise that retribution is properly a leading norm in our criminal justice system and presenting, in the form of modal retributivism, the best retributivist theory.¹¹⁷

Second, strictly speaking, modal retributivism is simply the theory that the culpable causing of a harm provides a reason to punish while culpable acting without causing harm provides merely a license to punish. The punishment levels recommended in the preceding discussion are not intrinsic features of modal retributivism. They and modal retributivism are jointly motivated

(discussing the atonement theory of punishment and comparing it to other popular theories, including restorativism); Huigens, *supra* note 91 (arguing in support of the aretaic theory of punishment, which has its roots in Aristotelian ethics).

115. Certain passages of Kant suggest that he thought that wrongdoers *must* be punished. According to Kant:

The law of punishment is a categorical imperative, and woe to him who crawls through the windings of eudaemonism in order to discover something that releases the criminal from punishment or even reduces its amount by the advantage it promises, in accordance with the pharisaical saying, "It is better for *one* man to die than for an entire people to perish." For if justice goes, there is no longer any value in human beings' living on the earth.

KANT, *supra* note 33, at 105. Compare Bedau, *supra* note 1, at 58 (describing as "fanatic" a person who ignores all costs of implementing retributivism), with Robinson, *supra* note 65, at 36-39 (1989) (providing a useful contemporary discussion of the trade-off between desert and utility).

116. MOORE, *supra* note 6, at 719-24. The view that deontological stricture make be overridden by extreme consequentialist concerns is called threshold deontology. See Moore, *supra* note 16, at 41 (summarizing literature on topic).

117. Others have advocated a similar position. See ALEXANDER & FERZAN, *supra* note 5, at 7-10 (defending "moderate" retributivism).

by the desideratum that moral truths be principled and nonarbitrary. But modal retributivism can be accepted even if these punishment levels are rejected.

Third, even if retributivism or the equality principle is rejected, modal retributivism may be relevant. A wide range of nonconsequentialist theories take harm to be at least a factor among those relevant to determining punishment. For example, a theory might recommend that where a harm is intentionally caused, the offender deserves a sanction of sufficient severity to communicate to the offender the wrongness of his conduct.¹¹⁸ With respect to actors who intentionally cause harm H , the necessary punishment might not be equal to H , but be a function of H , among other variables. With such theories, the issue will always arise: What is the appropriate amount of punishment in cases where (a) H was intended but not caused, and (b) H was merely risked but not caused? As with harm-based retributivism, making punishment a function of H in these cases, despite the fact that H was merely intended or risked, would be inconsistent with the intuition that harm matters, and making punishment a function of $R \times H$, or some other fraction of H , instead of H , would be problematic for the reasons given in Part II.c. The solution to this difficulty would be to make the sanction a function of H , or $R \times H$, respectively, changing the mode of punishment from precautionary to permissive. So, for example, under the theory that the purpose of punishment is to communicate to the wrongdoer the nature of her wrongdoing, the difference between wrongful acts that cause harm and those that do not could be reflected by the different mode of the sanction. The offender, who obviously must be given some information about why she is to be subjected to harsh treatment, could be informed it was either because of the harms she culpably caused, or her intent to cause harm plus some consequentialist justification, depending on whether harm was caused. This change in mode of punishment would communicate the significance of wrong done.¹¹⁹ Likewise, according to a theory in which the purpose of punishment is to assert the moral

118. See DUFF, *supra* note 4, at 351–53.

119. As a general matter, Duff, a proponent of the communicative theory of punishment, provides little guidance regarding levels of punishment. In his book on the law of attempts, he merely recommends that attempts receive a lighter sentence than the consummate offense correlate. *Id.* at 398.

value of the victim in the face of its denial by the wrongdoer,¹²⁰ where there has been no harm to the victim, and where there has been no mastery to refute¹²¹—the license to use the offender for the good of society may sufficiently reaffirm the value of the would-be victim. Accordingly, the approach of modal retributivism may be relevant to a wide range of nonconsequentialist theories of punishment.

VI. CONCLUSION

This article began with the widely held intuition that when it comes to punishment, results matter. It then posed the immediately obvious, but rarely asked, question, "How much do results matter?" In investigating this question, the article assumed that the principles of justice, be they compensatory, distributive, or retributivist, did not contain arbitrary constants or elements. A number of possible approaches to the problem were surveyed and found wanting on the ground they could not generate plausible, nonarbitrary penalty recommendations. Extended consideration was then given to the idea that where a person acts culpably, but there is no harm caused, the amount of punishment deserved should be calculated based on harm risked. The idea of punishing based on risk, however, was shown to be flawed because it could not be extended to a range of offenses, such as reckless endangerment, while remaining faithful to the intuition that results matter.

In light of these difficulties, modal retributivism was advanced. According to it, the fact that harm results from wrongful conduct is not relevant to the harshness of the sanction deserved, as it is under traditional harm-based retributivism, but rather, it is relevant to the sanction's mode as precatory ("should be imposed") or permissive ("may be imposed"). Allowing harm to determine the mode of punishment, rather than the amount, permitted harm to play a principled role in determining punishment, as well as offering a reasonable compromise between the claims of positive and negative retributivism. Finally, going beyond

120. See HAMPTON, *supra* note 21, at 130 ("[T]he *retributive* motive for inflicting suffering is to annul or counter the appearance of the wrongdoer's superiority and thus affirm the victim's real value.").

121. Hampton explains that where there is punishment, "I [the victim] master the purported master, showing that he is my peer." *Id.* at 125.

modal retributivism, this article sketched a principled approach to punishment for crimes of passion, as well as rejected punishment for crimes of negligence on the ground that a principled approach to them was not possible.

Even on the assumption that there are moral truths and principles of justice, establishing their content is no easy matter. Emotions and intuitions are our primary guides. But where emotions and intuitions are weak, ambiguous, vague, or indeterminate, reason can play a significant role in showing the way to moral truth, or at the very least, can develop and delimit a range of potential theories of justice that then might be evaluated for moral appeal. By introducing modal retributivism and highlighting its merits, this article hopefully has proven useful in this regard.
