

## RETRIBUTIVIST THEORIES' CONJOINED TWINS PROBLEMS

*Brittany L. Deitch\**

*This Article expands a previously published article, which introduced a novel problem to the centuries-old debate on the retributivist justification of punishment. The first article applied the problem of conjoined twins, where one commits a crime and the other is innocent, to pure retributivism. The conjoined twins problem showed that pure retributivism, which holds absolute duties to punish all who are guilty and none who are innocent, fails as a complete theory of punishment. This Article broadens the application of the conjoined twins problem by applying the problem to other versions of retributivism, including deontological, consequentialist, threshold, negative/weak, victim-conscious, and mixed retributivist theories. Exploring each version in turn, this Article uses the conjoined twins problem to show that no version of retributivism can serve as a complete theory of punishment.*

---

\* Case Writing Fellow, Harvard Law School. I would like to thank Russell Christopher and Stephen Galoob for piquing my interest in theories of punishment and for advising and supporting me throughout my first Article and this expansion. I would also like to thank Will Thomas for his valuable criticisms of an earlier version of this Article.

I. Introduction .....	140
II. Overview of Pure Retributivism's Conjoined Twins Problem .....	142
III. Impure Retributivism's Conjoined Twins Problems .....	145
A. Deontological .....	146
B. Consequentialist .....	148
C. Threshold .....	150
D. Negative/Weak .....	153
1. Negative/Weak: May We Punish This Actor? ....	154
2. Negative/Weak: To What Degree May We Punish This Actor? .....	157
E. Victim-Conscious .....	160
F. Mixed/Hybrid .....	164
IV. Conclusion .....	169

## I. INTRODUCTION

Consider the following:<sup>1</sup> A set of conjoined twins, *Guilty Twin* and *Innocent Twin*, has an argument with *Victim*. *Guilty Twin* murders *Victim* despite *Innocent Twin's* protests. *Guilty Twin* has complete dominion over the right arm used to wield the murder weapon. *Innocent Twin* could not prevent the crime from occurring. *Guilty Twin* is clearly guilty; *Innocent Twin* is clearly innocent. To avoid complicating this hypothetical with the legal and ethical questions surrounding a court-ordered separation for the twins, suppose the twins are conjoined in such a way that they share vital organs making separation impossible without causing the death of both twins. The twins stand trial with only *Guilty Twin* named as a defendant. If *Guilty Twin* is punished, *Innocent Twin* will necessarily suffer the punishment. However, forgoing punishing the innocent sets the guilty twin free. Thus, only two options are available: punish both twins or punish neither twin. What should the criminal law do?<sup>2</sup> What must<sup>3</sup> or may<sup>4</sup> a retributivist do?

---

1. This problem was originally presented in an earlier publication and applied to pure retributivism to demonstrate pure retributivism's failure to serve as a comprehensive theory of punishment. Brittany L. Deitch, *Retributivism's Conjoined Twins Problem*, 53 CRIM. L. BULL. 953 (2017).

2. Although this problem seems far-fetched or implausible, American courts have struggled with imposing criminal penalties upon conjoined defendants on at least four occasions. None of these cases addressed murder charges. The first involved assault. The second involved assault and battery. The third involved breach of the peace. The fourth involved an arrestable traffic offense. Deitch, *supra* note 1, at 964-66.

3. Strong versions of retributivism hold an absolute duty to punish. See, e.g., IMMANUEL KANT, THE METAPHYSICS OF MORALS 142 (Mary Gregor trans., 1991) (1797) (posing his famous disbanding island society hypothetical, and arguing that even though there would be no deterrent effect of punishment, every last murderer in that society's prison must be punished so that "blood guilt does not

Stated most simply and generally, retributivism is a theory that justifies punishment on the basis of the offender's culpability or desert.<sup>5</sup> *Retributivism's Conjoined Twins Problem*, this Article's predecessor, applied the conjoined twins problem to what may be termed pure retributivism, a "jealous theory"<sup>6</sup> that holds equal duties<sup>7</sup> and categorical imperatives to punish all who are guilty and none who are innocent.<sup>8</sup> That article discusses the conjoined twins problem with an emphasis on the significance of the personhood of the twins.<sup>9</sup> Regardless of whether a pure retributivist views the twins as two distinct persons or as two halves of one whole person, his theory fails to determine whether punishment of *Guilty Twin* is justified under the core tenets of his

---

cling to the people for not having insisted upon this punishment; for otherwise the people can be regarded as collaborators in this public violation of justice."); *See also*, MICHAEL MOORE, *PLACING BLAME: A THEORY OF THE CRIMINAL LAW* 154 (1997) ("As a theory of a kind of justice, [retributivism] obligates us to seek retribution through the punishment of the guilty. This means that officials have a duty to punish deserving offenders . . .") (emphasis in original); Russell L. Christopher, *Deterring Retributivism: The Injustice of "Just" Punishment*, 96 NW. U. L. REV. 843, 866 n.119 (2002) (explaining "obligatory retributivism," which holds that, because an offender has the "right to be punished," he must be punished) [hereinafter Christopher, *Deterring*].

4. Some versions of retributivism merely provide permission to punish. *See* Christopher, *Deterring*, *supra* note 3, at 866 n.120 (contrasting permissive retributivism with obligatory retributivism to explain that "permissive retributivism does not mandate punishment, but merely justifies its permissibility.").

5. *See* Christopher, *Deterring*, *supra* note 3, at 845 n.1 (2002) ("Retributivism is a theory, or justification, of punishment."); MOORE, *supra* note 3, at 92 ("[R]etributivism asserts that punishment is properly inflicted because, and only because, the person deserves it. That some people deserve punishment on such a theory is both a necessary and a sufficient condition justifying criminal sanctions."); Joel Feinberg, *The Classic Debate*, *THE PHILOSOPHY OF LAW* 614 (Joel Feinberg & Hyman Gross eds., 1991) ("Punishment is justified only on the ground that wrongdoing merits punishment."); A.M. Quinton, *On Punishment*, in *PHIL. PERSPECTIVES ON PUNISHMENT* 6, 7 (Gertrude Ezorsky ed., SUNY Press 1972) ("The essential contention of retributivism is that punishment is only justified by guilt."). For an overview of the debate between retributivism and its competing theory of punishment, consequentialism, *see generally* Christopher, *Deterring*, *supra* note 3, at 855-65.

6. *See* George P. Fletcher, *The Place of Victims in the Theory of Retribution*, 3 *BUFF. CRIM. L. REV.* 51, 52 (1999) ("[R]etributivism is a jealous theory in the sense that whatever the beneficial side-effects of punishment, if it is not deserved it cannot possibly be justified.").

7. *See* MOORE, *supra* note 3, at 157 ("[T]he guilty going unpunished is exactly the same magnitude of evil as the innocent being punished.").

8. The most prominent modern proponent of this theory is Michael Moore, who vigorously explains and defends pure retributivism in his book *Placing Blame*. *Id.* at 92 ("[R]etributivism asserts that punishment is properly inflicted because, and only because, the person deserves it. That some people deserve punishment on such a theory is both a necessary and a sufficient condition justifying criminal sanctions."); *Id.* at 154 ("As a theory of a kind of justice, [retributivism] obligates us to seek retribution through the punishment of the guilty. This means that officials have a duty to punish deserving offenders . . .") (emphasis in original); *see also* Michael T. Cahill, *Retributive Justice in the Real World*, 85 *WASH. U. L. REV.* 815, 848 (2007) (explaining that retributivists have equal moral duties to punish the guilty and not punish the innocent).

9. *See generally* Deitch, *supra* note 1, at 970 ("The retributivist might treat twins *A* and *B* as two halves of one whole person and elect to punish . . . On the other hand, if the retributivist treats conjoined twins as two distinct persons, then he must decide which duty supersedes the other.").

theory.<sup>10</sup> Under a two-person conceptualization, a pure retributivist would violate his positive duty to punish all who are guilty (with respect to *Guilty Twin*) if he forgoes punishing the twins, or his negative duty to never punish an innocent (with respect to *Innocent Twin*) if he opts to punish the twins. Under a same-person conceptualization, a pure retributivist would elect punishment of the twins because the twins are responsible for a murder. However, the same-person conceptualization is unsupported by legal and biological treatment of conjoined twins. Further, assigning desert-based punishment is insurmountably complicated when two minds—one guilty and one innocent—are at play. By exploring every possible approach, a pure retributivist might take in determining whether punishment is justified or prohibited, *Retributivism's Conjoined Twins Problem* concludes that pure retributivism fails as a comprehensive theory of punishment because of its inability to resolve the conjoined twins problem without violating one of its core tenets.

This Article presents and applies the conjoined twins problem to impure versions of retributivism. This expansion comprehensively explores all versions of retributivism to examine whether any existing variety of retributivism can satisfactorily resolve the conjoined twins problem. Part I succinctly summarizes *Retributivism's Conjoined Twins Problem* to establish the analytical framework employed throughout this Article. Part II provides brief descriptions of each version of retributivism: deontological, consequentialist, threshold, negative, victim-oriented, and mixed/hybrid. Following each overview is an application of that version's principles to the conjoined twins problem. After meticulously considering whether each version can satisfactorily resolve the conjoined twins problem under its tenets, this Article concludes that the conjoined twins problem reveals a fatal flaw of each and every desert-based justification of punishment. Even where a version seems able to resolve the problem through punishment or release of the twins, its resolution draws attention to the theory's other internal flaws.

## II. OVERVIEW OF PURE RETRIBUTIVISM'S CONJOINED TWINS PROBLEM

Pure retributivism holds two categorical, absolute duties: the positive duty to punish the guilty and the negative duty to never punish the innocent.<sup>11</sup> By contrast, the competing theory of punishment,

---

10. See generally Deitch, *supra* note 1, at 982.

11. See, e.g., Quinton, *supra* note 5, at 137 (“[W]e cannot punish the innocent and *must* only punish the guilty.”) (emphasis in original).

consequentialism, holds that punishment is justified when the good consequences for society flowing from the punishment outweigh the bad consequences for society.<sup>12</sup> Critics of pure retributivism argue that pure retributivism's twin duties cannot be satisfied in light of the common prosecutorial practices of plea-bargaining and immunity,<sup>13</sup> the practical impossibility of a zero-error rate in convictions,<sup>14</sup> and the inevitable suffering of the wrongdoer's family and friends.<sup>15</sup> Retributivists reply to these charges by rejecting the common prosecutorial practices,<sup>16</sup> relying upon the doctrine of double effect,<sup>17</sup> defining punishment to exclude the suffering of non-offenders from its definition,<sup>18</sup> or falling back on a softer version of retributivism.<sup>19</sup>

None of these standard defenses can save retributivism from the problem conjoined twins presents.<sup>20</sup> The conjoined twins problem does not depend on the existence of any particular practice or system of

---

12. See, e.g., Kent Greenawalt, *Punishment*, in 4 ENCYCLOPEDIA OF CRIME AND JUST. 1336, 1338 (Sanford H. Kadish ed., 1983) ("The dominant approaches to justification [of punishment] are retributive and utilitarian [a synonym for consequentialism]. Briefly stated, a retributivist claims that punishment is justified because people deserve it; a utilitarian believes that justification lies in the useful purposes that punishment serves.").

13. See Russell L. Christopher, *The Prosecutor's Dilemma: Bargains and Punishments*, 72 FORDHAM L. REV. 93 (2003-2004) (presenting the Prosecutor's Dilemma, where a prosecutor must violate one duty or the other when deciding whether to make an offer of immunity to one prisoner in order to punish that prisoner's confederates) [hereinafter Christopher, *Dilemma*]; Michael T. Cahill, *supra* note 8, at 854 ("Under the absolutist model [of retributivism], law enforcement at the prosecutorial stage would focus on making certain that all, and only, identified wrongdoers were punished in an amount appropriate to their desert. Accordingly, the absolutist model would categorically ban plea bargaining, witness immunity, downward departures for substantial assistance, or any other failure to impose deserved punishment.").

14. See, e.g., Richard O. Lempert, *An Assessment of the Moral Bases of the Case for Capital Punishment*, 79 MICH. L. REV. 1177, 1182 (1981) ("Retributivism is also haunted by those executions of the innocent which inevitably occur if the death penalty is allowed.").

15. See, e.g., Christopher, *Deterring*, *supra* note 3, at 879 ("Another way that retributivism is claimed to justify intentional punishment of the innocent is that infliction of punishment on a guilty offender will most likely inflict pain and suffering on his or her friends and family who are innocent of the offense.").

16. See Christopher, *Dilemma*, *supra* note 13, at 99 ("Retributivists implicitly assume that the incompatibility of bargain justice and retributivism warrants rejection of the former.").

17. See ST. THOMAS AQUINAS, SUMMA THEOLOGICA 1656 (Fathers of the English Dominican Prince trans., 1947) II-II, quest. 64, 7th art. (establishing the doctrine of double effect, which justifies an act which has both a good and a bad consequence if only the good consequence is intended and the bad consequence is an unfortunate side-effect); Christopher, *Deterring*, *supra* note 3, at 920 ("Since only the good effect of punishment of the guilty is intended and the bad side effect of punishment of the innocent or nonculpable is unintended, the conditions for [Aquinas' doctrine of double effect]'s applicability are satisfied.").

18. See Christopher, *Deterring*, *supra* note 3, at 879 ("[A] retributivist might reply that such suffering by friends and family members fails to fall within the notion of what is meant by punishment.").

19. See Deitch, *supra* note 1, at 963.

20. The remaining information in this section can be found in the earlier article. Deitch, *supra* note 1.

punishment, because it simply asks whether punishment is justified. The doctrine of double effect proves fruitless as a retributivist defense because punishment of *Innocent Twin* is not merely an unintended consequence of punishing *Guilty Twin*. Punishing *Guilty Twin* requires the retributivist to knowingly, intentionally punish an identifiable innocent, *Innocent Twin*. The definitional approach also fails, because if *Guilty Twin* is punished, *Innocent Twin* will suffer in exactly the same way as *Guilty Twin*. *Innocent Twin*'s suffering cannot fairly be categorized as anything but punishment because his suffering amounts to more than an incidental side-effect.<sup>21</sup> Retributivists' arsenal of existing defenses is thus unhelpful in avoiding the conjoined twins problem.

Turning to the problem of conjoined twins, the retributivist must answer two preliminary questions. First is an inquiry into whether the defendant is a person. Once the personhood inquiry is affirmatively established, the second step is to ask whether that person is a morally responsible agent. The first inquiry is usually taken for granted and has not been sufficiently explored in the past. While retributivist Michael Moore has developed a personhood theory for identifying whether punishment is justified, his theory is unhelpful in establishing the personhood status of conjoined twins. The second question usually arises in cases of insanity defenses or juvenile adjudications. After thoroughly considering sources from which retributivists might draw their personhood theory or definition, it becomes apparent that the retributivist ought to treat the conjoined twins as two persons. However, a particularly resolute retributivist could make some weaker arguments that the conjoined twins may be treated as one person.

Using a two-person conceptualization, pure retributivists must choose which core duty to honor and which to violate. To honor their positive duty to punish *Guilty Twin*, they must necessarily violate their negative duty to not punish *Innocent Twin*. To honor the negative duty by refusing to punish *Innocent Twin*, they must violate their positive duty to punish *Guilty Twin*. This is especially problematic for retributivists who claim that punishment is the "right" of the guilty, under the belief that punishment gives respect to the wrongdoer's autonomy.<sup>22</sup> It also

---

21. Deitch, *supra* note 1, at 963 (discussing criticisms that present examples of retributivists punishing innocents and retributivists' rebuttal that those examples do not amount to "punishment" in the technical sense, but merely reflect unfortunate side-effects of actual punishment).

22. See, e.g., Christopher, *Deterring*, *supra* note 3, at 864 ("Retributivism, however, honors offenders to such a degree that it conceives of punishment as the right of the offender."); G.W.F. HEGEL, ELEMENTS OF THE PHILOSOPHY OF RIGHT 126 (§100) (Allen W. Wood ed., H.B. Nisbet trans., 1991) (1821) ("The injury which is inflicted on the criminal is not only just *in itself* . . . it is also a *right for the criminal himself*"); HERBERT MORRIS, ON GUILT AND INNOCENCE: ESSAYS IN LEGAL PHILOSOPHY AND MORAL PSYCHOLOGY 41 (1976) ("a person has a right to be punished, meaning by this that a person has

presents a special problem for those retributivists who refuse to consider any person other than the offender in ascertaining whether the guilty deserves punishment. For these retributivists, accounting for *Innocent Twin's* suffering in determining whether *Guilty Twin* should be punished would force them to admit that retributivists should consider the suffering of the offender's family and friends, at least in some cases.

The same-person conceptualization allows retributivists to avoid the problem of choosing between their duties, but it presents a host of other problems. First, determining the level of culpability is difficult when there are two minds at play at the time of the offense—one guilty and one innocent. Second, intuition, biology, and the law all presuppose that the twins are two persons. Third, analogizing conjoined twins to a person with mental illness fails, because declaring conjoinedness a mental illness is unsupported by accepted psychology. Even if mental illness was a viable claim, this analogy would backfire because mentally ill persons are not considered morally responsible agents, resulting in reduced or no punishment.

Because pure retributivism centers on two absolutist duties, pure retributivists' analysis of whether they may justifiably punish the conjoined twins halts after the first two questions. Pure retributivists never reach the inquiries that follow the general decision that punishment is justified. In an ordinary case involving a non-conjoined wrongdoer, the pure retributivist would establish that the wrongdoer is (1) a person and (2) a morally responsible agent who is guilty and must be punished. Next, the pure retributivist would enter an analysis into the appropriate degree of punishment. Pure retributivists never reach this question in the conjoined twins problem because of their inability to handle the twins' personhood issues. Some impure versions of retributivism successfully advance past the personhood inquiries. However, as the analysis advances, they too suffer from fatal flaws in determining whether and to what degree the twins may justifiably be punished.

### III. IMPURE RETRIBUTIVISM'S CONJOINED TWINS PROBLEMS

Retributivism may generally be conceptualized as a desert-based<sup>23</sup>

---

a right to all those institutions and practices linked to punishment.”); JEFFRIE MURPHY, RETRIBUTION, JUSTICE, AND THERAPY 134 (1979) (“The right to be punished and regarded as a responsible agent, though sometimes painful when honored, at least leaves one’s status as a moral person in-tact.”); K.G. Armstrong, *The Retributivist Hits Back*, in THE PHILOSOPHY OF PUNISHMENT: A COLLECTION OF PAPERS 471, 484 (H.B. Acton ed., 1969) (“If we penalize the criminal according to what he has done, we at least treat him like a man, like a responsible moral agent.”).

23. See, e.g., Greenawalt, *supra* note 12, at 347 (“[P]unishment is justified because people deserve it.”).

and backward-looking<sup>24</sup> justification of punishment. Although this is the common thread of the retributivist theory, retributivism takes many forms.<sup>25</sup> Pure retributivism represents the strongest version of retributivism.<sup>26</sup> It is absolutist in its terms, firmly holding mandatory duties to always punish all who are guilty and to never punish any who are innocent. Many weaker versions of retributivism have emerged, retaining the desert-based and backward-looking methodology while avoiding the criticisms of pure retributivism.<sup>27</sup> This Part will provide a brief overview of the key tenets of the most prevalent impure versions of retributivism: deontological, consequentialist, threshold, negative/weak, victim-conscious, and mixed/hybrid.<sup>28</sup> An application of each theory to the conjoined twins problem follows each overview.

### A. Deontological

Deontological retributivism, like pure retributivism, views desert as both a necessary and sufficient justification of punishment.<sup>29</sup> Deontological retributivism focuses on the blameworthiness of each moral agent, considering each offender individually without accounting for other parties.<sup>30</sup> Because deontological retributivists evaluate the

---

24. See, e.g., A. Wesley Cragg, *Punishment*, in 2 THE PHILOSOPHY OF LAW: AN ENCYCLOPEDIA 605, 607 (Christopher Gray ed., Garland, 1999) (“[B]ackward-looking justifications see punishment as a response to moral wrongdoing.”); R.A. Duff, *Penal Communications: Recent Work in the Philosophy of Punishment*, 20 CRIME & JUSTICE 1, 6-7 (1996) (“[W]hat unites retributivist conceptions of punishment . . . is their insistence that punishment must be justified . . . in terms of its intrinsic character as a response to past wrongdoing.”).

25. See Christopher, *Deterring*, *supra* note 3, at 865 (“There are a considerable number of versions of retributivism.”).

26. See *supra* Part I.

27. See *supra* Part I.

28. Some versions of retributivism exist only to address particular aspects of justifying punishment that are not relevant to the inquiry the conjoined twins problem presents: Whether a retributivist may permissibly punish a guilty person when that punishment will also necessarily be knowingly imposed upon an identifiable innocent person. For example, moralistic retributivism holds that punishment is only justified when the offender’s desert is based upon a moral wrong, and legalistic retributivism holds that punishment is only justified when an offender’s desert is based upon his violation of a legal crime. Because the guilty twin in the conjoined twin problem has committed a murder, his conduct is unquestionably both immoral and illegal. Thus, he deserves punishment under both moralistic and legalistic retributivism. The question posed in the conjoined twins problem is an inquiry into whether the offender may be punished when an innocent would also be punished, not whether the conduct is eligible for punishment. For a discussion of moralistic and legalistic retributivism, see Christopher, *Deterring*, *supra* note 3, at 866.

29. See MOORE, *supra* note 3, at 156 (“The ‘deontological’ or ‘agent-relative’ retributivist regards the act of punishing the guilty as categorically demanded on each occasion, considered separately.”); Feinberg, *supra* note 5, at 614 (“In their pure formulations, [retributive theories] are totally free of utilitarian admixture. Moral or legal guilt . . . is not only a necessary condition for punishment, it is quite sufficient ‘irrespective of the consequences.’”).

30. MOORE, *supra* note 3, at 156 (“The ‘deontological’ or ‘agent-relative’ retributivist regards

desert of each individual, they would reject certain prosecutorial techniques that are designed to maximize punishment, such as plea bargaining.<sup>31</sup> For example, an offer of immunity to one offender in exchange for his testimony against his confederates would be prohibited under deontological retributivism, because the testifying offender must be punished for his own wrongdoing, even if such an offer would result in the conviction and punishment of his confederates.<sup>32</sup>

Deontological retributivists seeking to resolve the conjoined twins problem face approximately the same issues as pure retributivism. They hold absolute, obligatory duties to punish all who are guilty and none who are innocent. Because the conjoined twins problem requires the deontological retributivist to choose which duty to honor and which to violate, the deontological retributivist faces an irreconcilable dilemma. Because deontological retributivism is especially centered on treating each moral agent as an individual, without concern for others, the conjoined twins problem poses a greater threat to this version of retributivism, compared to the other impure versions discussed below. If the twins are treated as two persons, it is impossible for the deontological retributivist to determine whether punishment is justified. To honor his positive duty, the deontological retributivist must punish *Guilty Twin*. To honor his negative duty, he must not punish *Innocent Twin*.

The unique problem deontological retributivists face is the extraordinary difficulty of treating each twin as an independent individual. The twins, by their nature, are physically indivisible. To give *Guilty Twin* his just deserts would violate the negative duty by necessarily imposing the same punishment on *Innocent Twin*. To forgo punishment of *Innocent Twin* would necessarily release *Guilty Twin* from receiving his just deserts. The problem for a deontological retributivist is that his theory prohibits him from accounting for the desert of others in his determination of whether punishment is justified. The conjoined twins problem forces the deontological retributivist to consider the desert of one moral agent in assessing whether punishment

---

the act of punishing the guilty as categorically demanded on each occasion, considered separately.”). This probably traces its roots to Kant, who argued that every person must be treated as an end in himself. KANT, *supra* note 3, at 46-47 (“[E]very rational being, *exists* as an end in itself, *not merely as a means* to the discretionary use of this or that will.”; “*Act so that you use humanity, as much in your own person as in the person of every other, always at the same time as an end and never merely as means.*”) (emphasis in original).

31. See MOORE, *supra* note 3, at 158 (“[T]he deontological retributivist might simply deny the propriety of [common prosecutorial practices, like bargain justice].”).

32. *Id.* at 158 (“[T]he intentional forgoing of any opportunity to punish a guilty offender in order to obtain the conviction and punishment of [another] offender . . . is more of a problem for the deontological version of retributivism.”).

is justified for another moral agent. Thus, the mere acknowledgement of the difficulty he has in dealing with the conjoined twins violates deontological retributivism. To avoid this violation, the deontological could ignore the suffering of *Innocent Twin* and only look to the desert of the named defendant, *Guilty Twin*. However, once *Guilty Twin* is punished, *Innocent Twin*'s punishment is unjustified because it violates the negative duty of pure retributivism, which deontological retributivism adopts. The deontological retributivist cannot honor his duties to each twin and cannot fairly treat each twin as an individual.

Because deontological retributivism's application to the conjoined twins problem faces roughly the same problems as pure retributivism, with the additional problem of considering the desert of one in calculating the desert of another, deontological retributivists cannot resolve the conjoined twins problem.

### B. Consequentialist

Consequentialist retributivism is the first version of retributivism examined in this Article that softens the vengeful<sup>33</sup> tone of pure retributivism by drawing on consequentialism.<sup>34</sup> Consequentialist retributivists are motivated by retributivist principles of giving just deserts, but are willing to use consequentialist means to maximize desert-based punishment.<sup>35</sup> Perhaps consequentialist retributivism is best explained in contrast to deontological retributivism.<sup>36</sup> Deontological retributivists treat each person individually when determining whether to punish.<sup>37</sup> By contrast, consequentialist retributivists seek to maximize desert collectively.<sup>38</sup> Suppose a prosecutor is prosecuting three guilty persons.<sup>39</sup> If the prosecutor offers immunity to one of the guilty persons in exchange for his testimony against his two confederates, then the

---

33. Christopher, *Deterring*, *supra* note 3, at 848 ("From its roots in vengeance, bloodlust, revenge, retaliation, and eye for an eye, retributivism is pitched as the only theory which, in justifying punishment, does justice").

34. Christopher, *Dilemma*, *supra* note 13, at 146 ("The two alternative versions of retributivism – 'consequentialist' retributivism and threshold retributivism – attempt to incorporate the consequences of punishment into a retributivist framework.").

35. *Id.* at 147 ("[C]onsequentialist retributivism seeks to maximize the number of culpable wrongdoers punished.").

36. *Id.* at 146-47 ([Deontological retributivism] demands the punishment of each and every guilty person even if doing so foregoes greater opportunities to punish the guilty. In contrast, consequentialist retributivism seeks to maximize the number of culpable wrongdoers punished even if doing so foregoes punishment of some culpable wrongdoers.").

37. *Id.*

38. *Id.*

39. *Id.* at 123 (presenting the prosecutor's dilemma).

prosecutor will successfully convict the two.<sup>40</sup> However, by offering immunity to the one, the prosecutor is forgoing the opportunity to punish a guilty person.<sup>41</sup> If the prosecutor is a deontological retributivist, he will not offer immunity. An offer of immunity violates his duties to treat each person as an individual—without regard to others—and to punish all guilty persons.<sup>42</sup> A consequentialist retributivist, on the other hand, would offer immunity to that guilty person in order to punish the two.<sup>43</sup> The consequentialist retributivist will view this offer of immunity as having good, desert-based consequences.<sup>44</sup> By sacrificing one desert-based punishment, he will secure two desert-based punishments. The net gain of desert-based punishments justifies forgoing the punishment of the particular guilty actor.

As demonstrated by the bargain justice dilemma,<sup>45</sup> consequentialist retributivists are motivated by a desire to maximize desert. Consequentialist retributivists seek a net gain of desert-giving. In other words, forgoing punishment of one to convict and punish two is justified under consequentialist retributivism, but the same cannot be said of forgoing punishment of one to punish one. In the conjoined twins problem, *Innocent Twin* deserves no punishment and *Guilty Twin* deserves punishment. The consequentialist retributivist is only justified in deviating from stricter versions of retributivism if a net gain in desert-giving can be achieved from the deviation. Regardless of whether the consequentialist retributivist punishes or forgoes punishment, the result is a net-zero. Giving either twin his just desert necessarily results in a missed opportunity to give the other his just desert. Because there is no possibility for a net gain when only two persons with opposing deserts are involved, the necessary condition for deviating from the retributive framework is unsatisfied. Therefore, the consequentialist retributivist must rely on the baseline version of retributivism. Retributivism cannot resolve the conjoined twins problem because either result—punishing *Guilty Twin* or refusing to punish *Innocent Twin*—necessarily violates one of retributivism's core duties with respect to the other twin. Because

---

40. *Id.*

41. *Id.*

42. *Id.* at 118-22; See George Fletcher, BASIC CONCEPTS OF CRIMINAL LAW 36-37 (1998) (“It is patently unjust, in Kant’s view, to punish some offenders less because they are willing to cooperate in some way with the state.”).

43. See Christopher, *Dilemma*, *supra* note 13, at 149 (“Since punishing the Two maximizes the intrinsic good of retributive punishment, consequentialist retributivism provides a clear basis to accept the offer.”).

44. *Id.* at 147-48 (“Although both [deontological and consequentialist] versions of retributivism value a culpable wrongdoer receiving his just deserts, only consequentialist retributivism seeks to maximize generation of that intrinsic good.”).

45. See generally Christopher, *Dilemma*, *supra* note 13.

there is no net gain of desert-giving that could be achieved through either result, consequentialist retributivists cannot rely on their narrow incorporation of consequentialism to resolve the conjoined twins problem. Thus, consequentialist retributivism cannot resolve the conjoined twins problem any better than pure retributivism.

### C. Threshold

Threshold retributivism is the second version of retributivism that tempers its retributivist principles with consequentialism. Threshold retributivism, more commonly termed “threshold deontology,”<sup>46</sup> maintains the basic principles of deontological retributivism—that is, a morality-based and agent-based view of desert and punishment.<sup>47</sup> However, threshold retributivists limit the absolutist, Kantian view that “though the heavens may fall, justice shall be done.”<sup>48</sup> Under some circumstances, if the consequences are exceptionally bad, punishment of an innocent may be justified.<sup>49</sup> Critics of threshold retributivism claim that it collapses into consequentialism by considering the consequences in its analysis of whether punishment is justified.<sup>50</sup> Threshold retributivists combat this charge by arguing “[t]here is a very high threshold of bad consequences that must be threatened before something as awful as torturing an innocent person can be justified. Almost all real-life decisions . . . will not reach that threshold of horrendous consequences justifying torture of the innocent.”<sup>51</sup> Other critics demand to know why the consequences of punishment only matter at and above the threshold.<sup>52</sup> Threshold retributivists reply that the consequences

---

46. MOORE, *supra* note 3, at 719-24 (discussing the theory as “threshold deontology” throughout this section); Christopher, *Dilemma*, *supra* note 13, at 153 (“Threshold retributivism is merely threshold deontology applied to the punishment arena.”).

47. See MOORE, *supra* note 3, at 722 (describing threshold retributivism as “agent-relative views of morality with a threshold” and claiming that, “short of such a threshold, the agent-relative view [of deontological retributivism] will operate as absolutely as absolutism.”).

48. *Id.* at 719 (“[A]bsolutism is often attributed to Kant, who held that though the heavens may fall, justice must be done. Despite my non-consequentialist views on morality, I cannot accept the Kantian line.”).

49. *Id.* at 719 (For threshold retributivists, “It is just not true that one should allow a nuclear war rather than killing or torturing an innocent person. It is not even true that one should allow the destruction of a sizable city by a terrorist nuclear device rather than kill or torture an innocent person. To prevent such extraordinary harms extreme actions seem to me to be justified.”).

50. See Amartya Sen, *Rights and Agency*, 11 PHIL. AND PUBLIC AFFAIRS 3, 7 n. 8 (1982) (“Such a threshold-based ‘constraint’ system must rest ultimately on consequentialist analysis, comparing one set of consequences . . . with another . . . and its distinguishing feature will be the particular *form* of the consequence-evaluation function.”) (emphasis in original).

51. MOORE, *supra* note 3, at 722.

52. See Nancy (Ann) Davis, *Contemporary Deontology*, in A COMPANION ON ETHICS 205, 216 (Peter Singer ed., 1991) (“The addition of the ‘catastrophe clause’ is particularly problematic. Why do

“always count,” but they are insufficient to outweigh deontological concerns until a threshold is met.<sup>53</sup> Thus, threshold retributivists are essentially deontological retributivists who are willing to admit that sometimes there are good reasons to violate their core duties, either by punishing an innocent or by forgoing punishment of a guilty person.<sup>54</sup> Short of a consequence as catastrophic as a nuclear war resulting from honoring their duties, threshold retributivists rely on the principles of deontological and pure retributivism.<sup>55</sup>

A threshold retributivist would naturally start from the same position as a deontological retributivist in analyzing the conjoined twins problem. Unless a threshold-eligible catastrophe is identified, deontological retributivism guides the threshold retributivist.<sup>56</sup> After realizing that neither punishing *Guilty Twin* nor forgoing punishment of *Innocent Twin* can satisfy the core retributive duties, the threshold retributivist might be tempted to open his escape hatch and claim that the consequences of either decision are so deleterious that the other option is preferable. If he honors his positive duty and punishes *Guilty Twin*, the bad consequence would be the infliction of suffering upon *Innocent Twin*. If he honors his negative duty and forgoes punishing the *Innocent Twin*, then he would release a known murderer into society. Thus, the threshold retributivist might argue that the suffering of *Innocent Twin* is so bad that *Guilty Twin* must go free. He could alternatively argue that the consequence of letting a known murderer go free is so horrible for society that *Innocent Twin* may justifiably be punished. However, neither of these feared consequences, flowing from honoring either duty, rises to the level of catastrophe that threshold retributivism contemplates. Threshold retributivists are careful to limit the availability of the escape hatch to include only situations where there is “a very high threshold of bad consequences that must be

---

the effects of our acts on others' weal and woe acquire relevance only at the 'catastrophic' level?").

53. MOORE, *supra* note 3, at 723 (“[F]or a threshold deontologist consequences always ‘count.’ For behavior violating deontological constraints, however, until the threshold is reached, the [consequentialist principle] is outweighed by moral principles.”).

54. *Id.* at 723 (“As the consequences grow more and more severe, the consequentialist principle becomes of greater weight as applied to this situation, until at some point (the threshold) the consequentialist principle outweighs competing principles of morality.”); For a critical perspective, see Davis, *supra* note 51, at 215-16 (criticizing threshold retributivism and summarizing the theory as, “[A]llowing that we may violate deontological constraints in dire circumstances saves deontological views from the appearance of fanaticism, and thus confers greater normative plausibility on them.”).

55. See Christopher, *Dilemma*, *supra* note 13, at 156 (“Specific examples satisfying the threshold include the demise of a nation, if the heavens will in fact fall, nuclear war . . . the destruction of a sizeable city by a terrorist nuclear device.”).

56. *Id.* at 157 (explaining that, because the necessary threshold is not reached in the prosecutor's dilemma, threshold retributivists are bound by retributivism).

threatened.”<sup>57</sup> Short of that threshold, as with most real-life scenarios, “the agent-relative view . . . will operate as absolutely as absolutism in its ban on torturing the innocent.”<sup>58</sup>

The conjoined twins problem is a real-life scenario,<sup>59</sup> but it is a rare and exceptional problem. The frequency with which a problem occurs should have no bearing over whether a consequence meets the threshold; instead the focus should be on the magnitude of the bad consequence. However, even assuming *arguendo* that the rarity of the conjoined twins problem matters, the threshold retributivist still encounters difficulty in arguing that the negative consequence of *Innocent Twin*'s suffering or of *Guilty Twin*'s presence in society rise to the same high threshold as a nuclear war.<sup>60</sup> Evident in the examples of accepted thresholds—the demise of a nation, the fall of the heavens, the destruction of a sizable city, or the eruption of a nuclear war<sup>61</sup>—is an assumption that the bad consequence must be certain to occur and certain to affect a large population.<sup>62</sup>

If the threshold retributivist draws the line by accepting the suffering of one innocent as a reason not to punish one guilty, then he is left with no retributive theory of punishment at all. Instead, he collapses into consequentialism and loses his defense that the threshold must be “very high.” For example, if the threshold retributivist claims that releasing a known murderer into society satisfies the threshold, he would punish the twins. To reach this result, the threshold retributivist must argue that the consequences of forgoing punishment of a guilty person—without a showing that the guilty person will commit acts that are tantamount to the destruction of a sizeable city—can be used to justify the punishment of an innocent. This line of reasoning is flawed because one could apply it more broadly to justify nearly all punishment. Put simply, if the fear of releasing a guilty person on society can justify the intentional punishment of an innocent, then the fear of a person committing any future crime could justify the intentional punishment of innocents.

---

57. MOORE, *supra* note 3, at 722.

58. *Id.* at 722.

59. See Deitch, *supra* note 1, at 964 (“[C]onjoined twins have appeared as defendants in at least four American courts.”).

60. The analysis could change if *Guilty Twin* were released on this charge and proceeded to go on a killing spree. When he is prosecuted for the subsequent crimes and society's fears are realized and there is reason to believe that he will continue to wreak havoc, perhaps the punishment could then be justified under threshold retributivism. At that point, there is a possible argument that the consequences of refusing to punish *Innocent Twin* are tantamount to nuclear war.

61. See Christopher, *Dilemma*, *supra* note 13, at 156 (“Specific examples satisfying the threshold include the demise of a nation, if the heavens will in fact fall, nuclear war . . . the destruction of a sizeable city by a terrorist nuclear device.”).

62. See, e.g., MOORE, *supra* note 3, at 719 (expressing uncertainty about whether the destruction “of a lifeboat or building full of people” rises to the threshold).

This one-to-one ratio erring on the side of punishment could ultimately justify a system with a high rate of false convictions. Punishing more—for example, through adopting a lower standard of proof than beyond reasonable doubt—could be justified if innocents suffering is deemed necessary to avoid the prospective release of guilty persons. In addition to the problems of either favoring the positive duty over the negative or collapsing into consequentialism, this solution also rejects the Blackstone Principle, which Michael Moore indicates is a ratio that is compatible with his theory of threshold retributivism.<sup>63</sup>

Most threshold retributivists would likely concede that the feared consequences of honoring either duty do not meet the requisite threshold to justifiably violate a retributive duty.<sup>64</sup> Where the threshold is unmet, threshold retributivists may not use consequentialist considerations to resolve the problem.<sup>65</sup> Without license to use the escape hatch (consequentialism), the threshold retributivist is bound by deontological retributivism. As shown above, deontological retributivism provides no resolution for the conjoined twins problem. Thus, threshold retributivism fails to satisfactorily resolve the conjoined twins problem.

#### D. Negative/Weak

Scholars sometimes use the terms “negative” and “weak” retributivism interchangeably.<sup>66</sup> Negative or weak retributivism addresses two concerns in justifying punishment. The first is an inquiry into whether a particular actor may be punished. Negative retributivism responds by only adopting retributivism’s negative duty to never punish innocent persons.<sup>67</sup> The second concern asks how much punishment is

---

63. *Id.* at 157, n.11 (“The retributivist might adopt a principle of symmetry here – the guilty going unpunished is exactly the same magnitude of evil as the innocent being punished – and design his institutions accordingly. Or the retributivist might share the common view (that the second is a greater evil than the first) and design institutions so that ‘ten guilty persons go unpunished in order that one innocent not be punished.’”); *See generally* WILLIAM BLACKSTONE, 4 COMMENTARIES \*352. (“[I]t is better that ten guilty persons escape, than that one innocent suffer.”).

64. *See* Christopher, *Dilemma*, *supra* note 13, at 156-57 (arguing that the threshold is not met for threshold retributivists where “the torture of one innocent [prevents] the torture of the two [guilty offenders].”).

65. *Id.* at 157 (“Since a necessary condition for threshold retributivism is not met (the threshold), threshold retributivism fails to avoid the Prosecutor’s Dilemma.”).

66. *See, e.g.*, Christopher, *Deterring*, *supra* note 3, at 865 (“Weak, or negative, retributivism merely requires that a wrongdoer not be punished more than she deserves.”); *Id.* at 865.n.117 (“The term ‘weak retributivism’ may stem from Hart . . . The term ‘negative retributivism’ may derive from Mackie.”) (citing H.L.A. Hart, PUNISHMENT AND RESPONSIBILITY: ESSAYS IN THE PHILOSOPHY OF LAW 233 (Oxford Univ. Press 1992) (1968) and J.L. Mackie, *Morality and Retributive Emotions*, CRIM. JUST. ETHICS 3, 4 (1983)).

67. *See, e.g.*, Joshua Dressler, *Hating Criminals: How Can Something that Feels So Good Be Wrong?*, 88 MICH. L. REV. 1448, 1451 (1990) (book review) (“Negative retributivism . . . holds only

justified. Negative retributivism sets “an upper limit” on the permissible degree of punishment for guilty persons.<sup>68</sup> The following sub-sections treat these issues separately to illustrate the discrete issues the conjoined twins problem presents for negative retributivists.

### *1. Negative/Weak: May We Punish This Actor?*

Negative retributivism merely holds that “it is morally wrong to punish an innocent person.”<sup>69</sup> This may best be described in relation to pure retributivism. Recall that pure retributivism holds two duties—a positive duty to punish all who are guilty and a negative duty to never punish an innocent.<sup>70</sup> Negative retributivism may be viewed as a version of retributivism that adopts only the negative duty.<sup>71</sup> Another way to describe negative retributivism in relation to pure retributivism is to say that, where proponents of pure and deontological versions of retributivism hold criminal desert as a *sufficient* condition for punishment, negative retributivists claim only that desert is a “*necessary* condition of justified punishment.”<sup>72</sup>

Proponents of this version of retributivism argue that refusing to punish innocents is sufficient to “avoid the evil” of consequentialism.<sup>73</sup>

---

that it is morally wrong to punish an innocent person even if society might benefit from the action, i.e., the retributive principle of just deserts is a necessary condition of punishment.”); J.L. Mackie, *Morality and Retributive Emotions*, CRIM. JUST. ETHICS 3, 4 (1983) (defining negative retributivism as “the principle that one who is not guilty must not be punished”); JEFFRIE G. MURPHY, RETRIBUTIVISM AND THE STATE’S INTEREST IN PUNISHMENT 156, 159 (1970) (arguing that only the negative duty to not punish the innocent is necessary to avoid the evil— that is, taking rights of one person for a social good — that utilitarianism seems to invite) [hereinafter MURPHY, STATE]; R.A. Duff, *Penal Communications: Recent Work in the Philosophy of Punishment*, 20 CRIME & JUSTICE 1, 7 (1996) (“[A] ‘negative’ retributivism that forbids the punishment of the innocent (and perhaps also the excessive punishment of the guilty) has been more common . . . than a ‘positive’ retributivism that *demands* the punishment of the guilty, to the extent that they deserve.”) (emphasis in original).

68. Christopher, *Deterring*, *supra* note 3, at 869 (“Weak, or negative, retributivism merely requires that a wrongdoer not be punished more than she deserves.”); R.A. Duff, *Penal Communications: Recent Work in the Philosophy of Punishment*, 20 CRIME & JUSTICE 1, 7 (1996) (“[A] ‘negative’ retributivism that forbids the punishment of the innocent (and perhaps also the excessive punishment of the guilty) has been more common . . . than a ‘positive’ retributivism that *demands* the punishment of the guilty, to the extent that they deserve.”) (emphasis in original).

69. Dressler, *supra* note 66, at 1451.

70. See MOORE, *supra* note 3.

71. See J.L. Mackie, *Morality and Retributive Emotions*, CRIM. JUST. ETHICS 3, 4 (1983) (defining negative retributivism as “the principle that one who is not guilty must not be punished”); MURPHY, STATE, *supra* note 67, at 159 (arguing that only the negative duty to not punish the innocent is necessary to avoid the evil— that is, taking rights of one person for a social good — that utilitarianism seems to invite).

72. Duff, *supra* note 67, at 7 (emphasis in original); See also Dressler, *supra* note 66.

73. MURPHY, STATE, *supra* note 66, at 159 (arguing that only the negative duty to not punish the innocent is necessary to avoid the evil — that is, taking rights of one person for a social good — that utilitarianism seems to invite).

Retributivists frequently criticize consequentialism for its apparent ability to justify any heinous abridgement of an individual's rights, so long as the consequentialist can articulate a societal good derived from that abridgement.<sup>74</sup> By carving out a specific prohibition on punishing innocent persons and by refusing to adopt a requirement to punish guilty persons, negative retributivism offers arguably the least controversial, least restrictive theory of punishment. Most would agree that knowingly punishing innocent persons is morally wrong<sup>75</sup> and would likely feel less strongly about the morality of declining to punish a guilty person.

Critics of negative retributivism point out that it fails to qualify as a justification of punishment because it provides no affirmative reason to punish wrongdoers.<sup>76</sup> Negative retributivism articulates no guidelines for when punishment is justified. Instead, it merely states when punishment is unjustified by prohibiting punishment of innocents. If the purpose of justifying punishment is to set the terms for when it is morally acceptable to impose suffering on another,<sup>77</sup> then any theory of

---

74. See HART, *supra* note 66, at 5-6 (claiming that the argument that consequentialism can justify punishing the innocent is the "stock retributive argument" against consequentialism); Greenawalt, *supra* note 12, at 1341 ("[T]he most damaging aspect of the [retributivist] attack is that [consequentialism] admits the possibility of justified punishment of the innocent"); Christopher, *Deterring*, *supra* note 3, at 870 ("Retributivists' principal, and most devastating, criticism of consequentialist theories of punishment is that they justify punishment of the innocent.").

75. See generally WILLIAM BLACKSTONE, 4 COMMENTARIES \*352. ("It is better that ten guilty persons escape, than that one innocent suffer."); Keith A. Findley, *Toward a New Paradigm of Criminal Justice: How the Innocence Movement Merges Crime Control and Due Process*, 41 TEX. TECH. L. REV. 133, 136 (2008) ("Wrongful conviction of the innocent is a greater constitutional wrong than is failure to convict the guilty."); D. Michael Risinger, Essay, *Tragic Consequences of Deadly Dilemmas: A Response to Allen and Laudan*, 40 SETON HALL L. REV. 991, 1002 (2010) ("[F]or any given crime, an error that convicts an innocent person is much worse morally than an error that acquits a guilty person."); Hon. J. Harvie Wilkinson, Essay, *In Defense of American Criminal Justice*, 67 VAND. L. REV. 1099, 1109 (2014) ("[The Blackstone Ratio] is about balancing the twin aims of our criminal justice system: How do we punish as many of the deserving guilty as possible while ensnaring as few of the innocents as possible?"); Daniel Epps, *The Consequences of Error in Criminal Justice*, 128 HARV. L. REV. 1065, 1068 (2015) (defining the Blackstone principle as: "[I]n distributing criminal punishment, we must strongly err in favor of false negatives (failures to convict the guilty) in order to minimize false positives (convictions of the innocent), even if doing so significantly decreases overall accuracy."); see also *Genesis* 18:23-33 (Abraham asks God if He will destroy Sodom if fifty, forty-five, forty, thirty, twenty, or even ten righteous people can be found in the city. God replies, "For the sake of ten, I will not destroy it.").

76. See David Dolinko, *Some Thoughts About Retributivism*, 101 ETHICS 537, 539-44 (1991) (Negative retributivism provides no reason to punish a wrongdoer at all and does not serve as a complete theory of punishment.); R.A. Duff, *Penal Communications: Recent Work in the Philosophy of Punishment*, 20 CRIME & JUSTICE 1, 7 (1996) ("A merely negative retributivism, however, clearly provides no complete justification of punishment: for it tells us that we *may* punish the guilty (their punishment is not unjust), but not that or why we *should* punish them.").

77. See Cragg, *supra* note 24, at 707 ("One of the basic rules of civilized society is that deliberately inflicting pain and suffering on others always requires careful justification."); Greenawalt, *supra* note 12, at 1336 (outlining the three main concerns in justifying punishment as: (1) "why (and whether) the social institution of punishment is warranted," (2) what are "the necessary conditions for punishment in particular cases [,]" and (3) what "degree and severity . . . is appropriate for particular

punishment must articulate those terms. Put another way, defining something by what it is *not* offers no definition at all; likewise, a justification of punishment that fails to justify punishment is no justification of punishment.

Although negative retributivism might not support a penal system at all, one can conceive of a version of negative retributivism that adopts the negative duty to never punish an innocent as obligatory and adopts the positive duty to punish the guilty as merely permissive (and thus not as a technical “duty”).<sup>78</sup> If a negative retributivist supports a penal system under these guidelines, he is vulnerable to the same criticisms leveled against pure retributivism with respect to its failure to satisfy its negative duty.<sup>79</sup> In other words, every argument claiming pure retributivism violates its negative duty by supporting or endorsing a system that inevitably produces erroneous guilty verdicts<sup>80</sup> and causes the offenders’ innocent loved ones to suffer<sup>81</sup> also applies to this conceptual version of negative retributivism.

Negative retributivism appears well-suited to resolve the conjoined twins problem. By only adopting the negative duty, a negative retributivist is unplagued by a need to honor the competing positive duty. *Guilty Twin* deserves punishment, but the negative retributivist bears no obligation to punish him. *Innocent Twin* deserves no punishment, and the negative retributivist is obligated to set him free. The solution for the negative retributivist seems clear: satisfy the only obligatory duty by forgoing punishment of the twins.

Because the negative retributivist need not choose between competing duties, the conjoined twins problem does not create a moral dilemma<sup>82</sup>

---

offenses and offenders.”).

78. No scholar has yet to endorse such a theory, but it is logically possible and will be considered in this Article for the sake of comprehensiveness. Arguably the theory presented here is part of H.L.A. Hart’s theory discussed in Part II E.

79. See generally Deitch, *supra* note 1, at 961 (discussing criticisms that retributivists violate their negative duty when they endorse a system of punishment).

80. See Richard O. Lempert, *Desert and Deterrence: An Assessment of the Moral Bases of the Case for Capital Punishment*, 79 MICH. L. REV. 1177, 1182 (1981) (“Retributivism is also haunted by those executions of the innocent which inevitably occur if the death penalty is allowed.”); Christopher, *Deterrence*, *supra* note 3, at 900 (“Retributivists readily acknowledge that innocents will mistakenly be punished, but dismiss it as an insignificant problem. Retributivists claim that though they know that some unknown innocents will be mistakenly punished, no particular, identifiable innocent is intentionally punished under retributivism.”).

81. See Christopher, *Deterrence*, *supra* note 3, at 879 (“Another way that retributivism is claimed to justify intentional punishment of the innocent is that infliction of punishment on a guilty offender will most likely inflict pain and suffering on his or her friends and family who are innocent of the offense.”); Richard O. Lempert, *Desert and Deterrence: An Assessment of the Moral Bases of the Case for Capital Punishment*, 79 MICH. L. REV. 1177, 1182 (1981) (“Retributivism is also haunted by those executions of the innocent which inevitably occur if the death penalty is allowed.”).

82. See Christopher, *Dilemma*, *supra* note 13, at 137-42 (defining moral dilemma generally as a situation in which an actor must select an option that violates his principles in order to satisfy another

for the negative retributivist. However, the conjoined twins problem illustrates and illuminates negative retributivism's other flaws. By releasing *Guilty Twin* based upon the absence of a duty to punish him, the negative retributivist offers *Guilty Twin* a biologically built-in, limitless defense to any crime. As long as *Innocent Twin* remains innocent through *Guilty Twin*'s commission of offenses, *Guilty Twin* never receives his just desert. Notice though that *Innocent Twin*'s innocence is probably irrelevant with respect to negative retributivism's terms. By holding only the negative duty to forgo punishing innocents, the negative retributivist may permissibly never punish *anyone*, regardless of the person's guilt or physical connectedness to another.

Retributivism, at its core, is a desert-based justification for punishment. If this version of retributivism can justify indefinitely forgoing punishment of a known murderer for no apparent reasons with no internal qualms, then it cannot truly be a version of retributivism. It neither issues punishment in accordance with desert nor justifies punishment at all. Negative retributivism can, at best, be viewed as a limitation on punishment. A theory of punishment that knowingly allows a person who is guilty of a serious offense to go unpunished simply fails as a desert-based justification of punishment. Because negative retributivists can easily resolve the conjoined twins problem by forgoing punishment, and because the physical nature of the twins plays no role in the analysis, negative retributivism is neither retributive nor a justification of punishment.

## 2. *Negative/Weak: To What Degree May We Punish This Actor?*

The above description of negative retributivism addresses the question of whether punishment is justified (though, as explained, it only addresses the question of whether punishment is prohibited – and may feasibly permit always declining to punish). Another explanation of negative retributivism deals with the degree of permissible punishment for a guilty actor.<sup>83</sup> Here, negative retributivism “sets the upper limit” on the amount of punishment permissibly imposed upon a guilty offender.<sup>84</sup>

The degree-based form of negative retributivism establishes a cap on the amount of punishment that may justifiably be imposed upon a person.<sup>85</sup> A guilty actor must not be punished more than he deserves.<sup>86</sup>

---

principle, and discussing the various types of moral dilemmas on a spectrum from broad to narrow).

83. See Christopher, *Deterring*, *supra* note 3, at 869 (discussing Hart's mixed theory's assignment of amount of punishment and explaining, “The negative retributivism component [of Hart's theory] sets the upper limit . . . on how much punishment to impose.”).

84. *Id.*

85. *Id.*

Naturally, this also forbids punishing an innocent person, because an innocent deserves precisely no punishment. Like each version of retributivism discussed thus far, which prohibit punishment of innocents, punishment of an innocent is also categorically prohibited under this version.

All versions of retributivism hold that punishing an innocent is unjustified,<sup>87</sup> but retributivists disagree about how much punishment a guilty person deserves.<sup>88</sup> Pure retributivism, and any version of retributivism adopting the strong, positive duty, holds that guilty persons must be punished to the fullest extent of their desert.<sup>89</sup> Negative retributivists do not adopt this requirement.

Degree-based negative retributivism is generally used as a guiding principle in mixed theories and is not employed as a stand-alone justification of punishment. For this reason, it is not subject to every criticism leveled against the form of negative retributivism that seeks to answer the question of whether punishment is permissible.<sup>90</sup> However,

---

86. *Id.* at 865 (“Weak, or negative, retributivism merely requires that a wrongdoer must not be punished more than she deserves.”).

87. *See, e.g.*, Talia Fisher, *Conviction Without Conviction*, 96 MINN. L. REV. 833, 875-76 (2012) (“Retributivists are concurrently committed to two fundamental principles: punishing the guilty and not punishing the innocent. Deviation from either one of these outcomes is considered a departure from the principles of just desert.”); *But see infra* Part II. C. (explaining that threshold retributivists admit that, under extreme circumstances, intentional punishment of the innocent may be justified).

88. *See* Larry Alexander, *You Got What You Deserved*, 7 CRIM. L. & PHIL. 309, 315 (2012) (“The ‘how much is deserved’ question has many facets, some of which have been copiously debated.”). Common views of proportionality among retributivists are Kant’s *lex talionis* theory, Hegel’s annulment theory, and Jean Hampton’s expressive theory. KANT, *supra* note 3, at 169 (equality between the crime and the punishment, which is also called *lex talionis* or “an eye for an eye”); HEGEL, *supra* note 22, at 127 (establishing the principle that the offender’s suffering should be proportionate to the value of the crime and that “what the criminal has done should also happen to him.”); Jean Hampton, *Correcting Wrongs Versus Righting Wrongs: The Goal of Retribution*, 39 UCLA L. REV. 1659, 1686 (1992) (setting forth the “expressive” theory of retribution and advocating for equality between the suffering of the criminal and the suffering he caused the victim, which demands that “a response to a wrong that is intended to vindicate the value of the victim denied by the wrongdoer’s action through the construction of an event that not only repudiates the action’s message of superiority over the victim but does so in a way that confirms them as equal by virtue of their humanity.”); Adam J. Kolber, *The Comparative Nature of Punishment*, 89 B.U. L. REV. 1565, 1595 (2009) (“Moreover, if we want to punish proportionally, then we have to calibrate punishments to reflect the suffering that offenders actually experience or are expected to experience as a result of being punished.”).

89. *See* KANT, *supra* note 3, at 169 (punishment requires *lex talionis*, equality between the crime and the punishment); HEGEL, *supra* note 22, at 127 (“[W]hat the criminal has done should also happen to him.”); Christopher, *Deterring*, *supra* note 3, at 865-66 (“[S]trong, or positive, retributivism requires that a wrongdoer be punished to the fullest extent of his just deserts.”).

90. Stated differently, the form of negative retributivism that attempts to answer the question of whether punishment is permissible can be criticized for failing to provide any reason to answer that question in the affirmative. By contrast, the desert-based form of retributivism does not attempt to justify punishment or explain when punishment is unjustified. It merely addresses the question of how much punishment a person should get after using some other theory to determine that punishment is justified.

degree-based negative retributivism is vulnerable to some criticisms leveled against the other form of negative retributivism, including the inevitability of a zero-error rate in any criminal justice system and the suffering of the guilty actor's innocent loved ones. For the latter criticism, a degree-based negative retributivist could conceivably reply that considering the suffering of others in calculating the appropriate degree of punishment for the guilty actor is permissible. Given the limited nature of the question that degree-based negative retributivism answers, there is nothing in the theory that addresses whether only the guilty actor may be considered in the degree calculation. Because degree-based negative retributivism is folded into other theories, it could conceivably be combined with a version of consequentialism that accounts for a guilty actor's familial status or role in the community.

Additionally, degree-based negative retributivism receives its own unique, proportionality-based criticisms. For example, critics have argued that mixed theorists who use degree-based negative retributivism are "disproportionately lenient" because they decline to "[justify] the punishment of guilty wrongdoers to the fullest extent of their just deserts."<sup>91</sup> A degree-based negative retributivist might reply that it is morally acceptable to fail to punish to the fullest extent permitted because preventing excessive punishment and refusing to punish innocents is all that is morally required to avoid evil.<sup>92</sup>

Turning to the conjoined twins problem, assume the negative retributivist punishes. He will then need to determine the degree of punishment. *Guilty Twin* deserves punishment to a high degree, because he is guilty of a murder. However, *Innocent Twin* deserves no punishment at all, because he is innocent. The negative retributivist may not impose punishment in excess of the person's desert. One person—*Guilty Twin*—deserves a high level of punishment and the other—*Innocent Twin*—deserves absolutely none, but any punishment will necessarily be inflicted equally upon both. There is no way to reconcile the problem presented. A negative retributivist may punish *Guilty Twin* to any degree up to what he deserves, so the negative retributivist might impose a punishment less than the fullest extent permissible. Whatever punishment the negative retributivist inflicts upon *Guilty Twin* will also be imposed upon *Innocent Twin*. With respect to *Innocent Twin*, no amount of punishment can be justified under negative retributivism because any punishment exceeds his desert. Thus, any punishment of *Guilty Twin* would violate negative retributivism's duty to *Innocent*

---

91. Christopher, *Deterring*, *supra* note 3, at 869.

92. See MURPHY, *STATE*, *supra* note 66, at 159 (arguing that only the negative duty to not punish the innocent is necessary to avoid the evil— that is, taking rights of one person for a social good — that utilitarianism seems to invite).

*Twin.*

Now, assume the negative retributivist forgoes punishing *Guilty Twin*. Failure to punish is permissible with respect to both *Innocent Twin* and *Guilty Twin*. The problem here is largely the same as the problem in the preceding section. Failure to punish *Guilty Twin* is permissible under the theory, but it is also extraordinarily lenient. Releasing a known murderer is inconsistent with retributivism, a desert-based justification of punishment. Apart from its inability to adhere to the retributive framework, negative retributivism's release of a known murderer for no apparent reason demonstrates its failure to serve as a justification of punishment at all.

### E. Victim-Conscious

Some retributivists argue that victims should play a role in justifying retributive punishment.<sup>93</sup> Whereas consequentialists might consider both actual and putative victims in justifying punishment,<sup>94</sup> victim-conscious retributivists—as advocates of a backward-looking theory of punishment<sup>95</sup>—carefully limit their consideration only to actual victims of the specific past crime for which the offender is guilty.<sup>96</sup>

Victim-conscious retributivist, George Fletcher, analogizes his theory

---

93. See Fletcher, *Victims*, *supra* note 6, at 63 (“[T]he interests of victims may be properly integrated into the theory of retributive punishment.”).

This traces back to Aristotle’s Theory of Justice, wherein Aristotle argues for justice as rectification:

“It makes no difference whether a decent man has defrauded a bad man or vice versa, or whether it was a decent or bad man who committed adultery. The only difference the law considers is that brought about by the damage: it treats the parties as equal and asks only whether one has done and the other has suffered wrong, and whether one has done and the other has suffered damage. As the unjust in this sense is inequality, the judge tries to restore equilibrium. When one man has inflicted and another received a wound, or when one man has killed and the other has been killed, the doing and suffering are unequally divided; by inflicting a loss on the offender, the judge tries to take away his gain and restore the equilibrium.”

ARISTOTLE, *NICOMACHEAN ETHICS* 120-21 (Martin Oswald trans., 1969) (350 B.C.E.). Fletcher, *Victims*, *supra* note 6 at 58 (“Seeking to bring about equality between victim and offender is a classic concern in Aristotle’s theory of justice.”).

94. See Adam J. MacLeod, *All for One: A Review of Victim-Centric Justifications for Criminal Punishment*, 13 BERKELEY J. CRIM. L. 31, 37 (2008) (“[Peter Singer, a consequentialist] does not substitute traditional consequentialist concerns . . . for guilty usurpation of society’s legal standards as a rationale or punishment. He instead uses impediment to the putative victim’s legal interest.”) (citing Peter Singer, *Pulling Back the Curtain on the Mercy Killing of Newborns*, L.A. TIMES Mar. 11, 2005, at B13).

95. See, e.g., Cragg, *supra* note 24, at 707 (“[B]ackward-looking justifications see punishment as a response to moral wrongdoing.”)

96. See Fletcher, *Victims*, *supra* note 6, at 55 (“First, the victims that are relevant for our purposes are the actual victims not the potential victims of future crimes.”).

to Hegel's norm vindication theory.<sup>97</sup> Hegel argues that retributive punishment seeks to restore the norm by defeating the aggressor proportionate to the amount in which he disrupted the norm.<sup>98</sup> He famously characterizes punishment as "the negation of the negation."<sup>99</sup> For Hegel, the norm is defended by the punishment,<sup>100</sup> but for Fletcher, "the 'victim' takes the place of the 'norm' in the structure of the argument."<sup>101</sup> While Hegel advocates for punishment that negates the negation the wrongdoer created for the state generally, Fletcher views the aim of punishment as restoring equality between the wrongdoer and the victim.<sup>102</sup> When a wrongdoer does wrong, according to Hegel, he gains an unfair advantage over society that must be corrected to restore equality. For Fletcher, when a wrongdoer does wrong, he gains unfair dominance over his victim and equality must be restored.<sup>103</sup> Fletcher incorporates victims into both the corrective<sup>104</sup> and the distributive<sup>105</sup> dimensions of punishment.<sup>106</sup> The corrective dimension is satisfied when an offender is punished and the punishment "communicat[es] to the victim a concern for his or her antecedent suffering."<sup>107</sup> Through victim-conscious corrective justice, the state punishes offenders to express solidarity with victims. The distributive dimension is satisfied when offenders are punished equally with respect to one another.<sup>108</sup>

---

97. *Id.* at 58 ("This victim-based argument does not differ, in principle, from the Hegelian argument that punishment serves to vindicate the norms against those who have sought to defeat it. The only difference is that the 'victim' takes the place of the 'norm' in the structure of the argument."); Hampton, *supra* note 88, at 1694 ("As Hegel appreciated, the modern state is the citizenry's moral representative; in the face of pluralism and religious controversy, it is the only institutional voice of the community's shared moral values.").

98. See HEGEL, *supra* note 22, at 127 ("[W]hat the criminal has done should also happen to him.").

99. *Id.* at 123 ("[P]unishment is merely the negation of the negation.").

100. See Fletcher, *Victims*, *supra* note 6, at 58 ("In the traditional Hegelian view, the norm is defended and the aggressor symbolically defeated.").

101. *Id.*

102. *Id.* ("[T]he position and dignity of the victim are rendered equal relative to the aggressor.").

103. *Id.* ("The function of arrest, trial, and punishment is to overcome this dominance and reestablish the equality of victim and offender.").

104. Corrective justice can be characterized as equality between the victim and the offender. *Id.* at 58 ("Corrective justice seeks the equality that existed between the victim and offender prior to the wrongful act.").

105. Distributive justice can be characterized as equality between and among like offenders. *Id.* at 58 ("The distributive dimension of punishments consists in the legal imperative to punish all offenders equally. The evil of punishment should be distributed fairly, with each offender receiving his just deserts."); Cf. MOORE, *supra* note 3, at 87-88 ("Punishment of the guilty is thus for the retributivist an *intrinsic* good, not the merely *instrumental* good that it may be to the utilitarian or rehabilitative theorist.").

106. See Fletcher, *Victims*, *supra* note 6, at 58 ("Retributive justice combines features of both corrective and distributive justice.").

107. *Id.* at 58.

108. *Id.* ("The distributive dimension of punishment consists in the legal imperative to punish all

The obvious criticism of victim-conscious retribution is that it fails to justify punishment for crimes in which there is no identifiable victim.<sup>109</sup> In response, Fletcher “fills the gap”<sup>110</sup> by drawing on the retributive, Kantian<sup>111</sup> desire to avoid “the evil of *impunidad*—the phenomenon of offenders getting away with their crimes.”<sup>112</sup> By adding this component to his theory, Fletcher limits his victim-conscious rationale to justify crimes in which there is an identifiable victim, but maintains the Hegelian rationale to justify crimes that merely unfairly elevate the wrongdoer above society. Although Fletcher has attempted to justify his victim-conscious approach using Hegel’s and Kant’s retributive reasoning, critics nevertheless maintain that the underlying “power-imbalance calculus” is inherently consequentialist.<sup>113</sup> By justifying punishment with the focus on the balance of society as a whole, instead of merely assessing the offender’s desert, Fletcher’s victim-conscious version of retributivism collapses into consequentialism.

Jean Hampton takes a slightly different approach in her victim-conscious, “expressive theory of retribution.”<sup>114</sup> Rather than viewing completion of wrongful conduct as the offender gaining dominance or an unfair advantage over society, Hampton focuses on “moral injury,” defined as “an affront to the victim’s value or dignity.”<sup>115</sup> The purpose of the expressive theory of retribution is to correct the injury by acknowledging and restoring the victim’s value. Punishment, she argues, is the appropriate mechanism by which to elevate the value of the victim.<sup>116</sup> Unlike a parade for the victim, for example, punishment creates an actual state of affairs.<sup>117</sup> By analogizing retribution to

---

offenders equally.”).

109. See MacLeod, *supra* note 94, at 42 (“Fletcher’s conception of punishment as a means to restore balance between victim and offender is criticized as inconsistent with retributivism . . . Fletcher acknowledges some of the difficulties . . . for example, that many crimes involve no dominance over a readily-identifiable victim.”).

110. *Id.* at 42–43 (“Avoidance of *impunidad*, he argues, fills the gap left by his dominated-victim theory.”).

111. KANT, *supra* note 3.

112. Fletcher, *Victims*, *supra* note 6, at 60.

113. MacLeod, *supra* note 93, at 45 (“[Fletcher] has moved from the ranks of retributivists to the ranks of consequentialists . . . The community is justified, in Fletcher’s view, in punishing a person in order to restore an inherently consequentialist power-imbalance calculus. If this is what Fletcher intends to argue, then . . . Fletcher has surrendered his membership in the retributivist club.”).

114. Hampton, *supra* note 88, at 1659.

115. *Id.* at 1666.

116. *Id.* at 1695 (“[P]unishment of the wrongdoer is uniquely suited to do the vindicating, insofar as it created a state of affairs . . . in which the victim was elevated with respect to the wrongdoer.”).

117. *Id.* (describing a critic’s inquiry into why we should do harm to the wrongdoer instead of bestowing a benefit onto or throwing a parade for the victim, Hampton replies, “[P]unishment of the wrongdoer is uniquely suited to do the vindicating, insofar as it created a state of affairs . . . in which the victim was elevated with respect to the wrongdoer.”).

damages in torts claims, Hampton argues that retributive punishment compensates the victim for his or her moral injuries.<sup>118</sup> To emphasize the importance of the victim, Hampton also argues that the victim must be someone other than the wrongdoer himself, rejecting a retributive response to someone who attempts suicide.<sup>119</sup>

What Fletcher and Hampton both support is a version of retributivism that stands in solidarity with victims. Where they differ is in their narrower focus. For Fletcher, the wrongdoer gains an advantage over his victim. For Hampton, the victim suffers an injury at the hands of the wrongdoer. Under both theories, the state must correct an imbalance by giving a wrongdoer his just deserts.

It might be helpful to apply victim-conscious retributivism to the conjoined twins problem by altering the problem slightly. Suppose that instead of murdering *Victim*, *Guilty Twin* inflicts physical injury upon *Innocent Twin*.<sup>120</sup> *Guilty Twin* is clearly guilty and deserves punishment. *Innocent Twin* is not only innocent but is also the very victim of the offense. Punishment imposed upon *Guilty Twin* will necessarily be imposed upon *Innocent Twin-Victim*. Victim-conscious retributivists justify punishment of wrongdoers by viewing punishment as restoring equality between the victim and the offender or as vindicating the value of the victim. These theorists adopt punishment as the appropriate mechanism to restore equality. However, in this scenario, punishment will cause further harm to the victim. Equality may never be restored because, so long as *Guilty Twin* is punished, *Innocent Twin-Victim* is equally punished.

It may help to frame the theories in mathematical terms. Under Fletcher's theory, the offense puts *Guilty Twin* at a value of positive one (1) and leaves *Innocent Twin-Victim* at net-zero (0). Punishment should be imposed upon *Guilty Twin* to remove his unfair gain and restore equality at 0. However, if punishment counts as negative (-1) with respect to *Guilty Twin*, it also must count as negative one (-1) with respect to *Innocent Twin-Victim*. Thus, *Guilty Twin* may be brought down to net-zero (0), but *Innocent Twin-Victim* will be reduced to negative one (-1). Equality between the offender and the victim cannot be restored by using victim-conscious, corrective justice.

Under Hampton's theory, the offense causes *Innocent Twin-Victim* to

---

118. See Hampton, *supra* note 87, at 1698 (“[R]etribution is actually a form of compensation to the victim . . . retribution is supposed to . . . [place] the victim in the position she would have been in had the wrongdoer not acted.”).

119. *Id.* at 1665 (“We also think it deeply inappropriate to inflict a retributive response on someone who has wronged himself.”).

120. This alteration was also presented in *Retributivism's Conjoined Twins Problem*. Deitch, *supra* note 1, at 977.

decrease in value from a net-zero (0) to a negative one (-1). Hampton's theory requires *Guilty Twin's* punishment to restore *Innocent Twin-Victim's* value to net-zero (0). But punishing *Innocent Twin-Victim* constitutes a moral injury to *Innocent Twin-Victim* by violating *Innocent Twin-Victim's* dignity. Because *Innocent Twin-Victim* does not deserve punishment, his dignity is further reduced by *Guilty Twin's* punishment. The state's punishment of *Guilty Twin* may restore *Innocent Twin's* value to net-zero (0), but the state's punishment of *Innocent Twin-Victim* reduces *Innocent Twin-Victim's* value down to negative one (-1). Under Hampton's theory, punishment fails to achieve its objective because the numeric value of the offender and the victim remain the same. It is also worth noting that, because Hampton rejects retributive punishment as a response to self-harm, a same-person conceptualization of the conjoined twins results in forgoing punishment of the twins. This forgone punishment constitutes excessive leniency with respect to *Guilty Twin*, giving *Guilty Twin* a biologically built-in defense to commit further crimes, and fails to restore *Innocent Twin's* value.

Victim-conscious versions of retributivism cannot justify punishment of *Guilty Twin*, because the punishment will necessarily also be inflicted on *Innocent Twin-Victim*. Equality cannot be restored, and the victim's dignity cannot be acknowledged under either result. If the victim suffers the same punishment as his aggressor, he will not be restored or acknowledged. If punishment is forgone, the victim will not be made to suffer through punishment, but neither will the offender. Under either result, victim-conscious retributivism fails to achieve its primary goal: solidarity with the victim. Thus, victim-conscious retributivism cannot satisfy the conjoined twins problem.

#### F. Mixed/Hybrid

In an apparent effort to avoid the criticisms of either broad justification of punishment—consequentialism and retributivism – many scholars have devised mixed or hybrid theories incorporating aspects of both broad theories.<sup>121</sup> Because of the volume mixed/hybrid versions of retributivism,<sup>122</sup> it is not possible to include all of them in this Article.

---

121. See Christopher, *Deterring*, *supra* note 3, at 867 (“In response to the seemingly intractable debate between retributivist and consequentialist theories of punishment, there have been attempts to combine the two.”).

122. Mixed theories have been offered since Cesare Beccaria presented his theory in 1764, which combines utility (consequentialism) with justice (retributivism). See CESARE BECCARIA, ON CRIMES AND PUNISHMENTS (David Young trans., 1986) (1764). For a brief overview of Beccaria's theory, see Christopher, *Deterring*, *supra* note 3, at 867, n. 128. For an overview of the various mixed theories, see IGOR PRIMORATZ, *Mixed Rationales*, in 2 THE PHILOSOPHY OF LAW: AN ENCYCLOPEDIA 559-61 (Christopher Gray ed., Garland, 1999).

Instead, this Article discusses two: the most prominent version of mixed retributivism and one of the most recent versions.

H.L.A. Hart presented his theory in 1968,<sup>123</sup> and it remains the most prominent mixed theory of punishment.<sup>124</sup> Hart separates punishment justification issues into two categories: the General Justifying Aim and Distribution.<sup>125</sup> The General Justifying Aim inquiry seeks to justify “the practice of a system of punishment.”<sup>126</sup> The Distribution inquiry justifies “who may be punished.”<sup>127</sup> Finally, Hart’s theory also presents a proportionality component within the Distribution analysis that justifies the amount of punishment imposed.<sup>128</sup>

For Hart, the General Justifying Aim of Punishment is reasoned supported by consequentialism.<sup>129</sup> The means used in individual punishments to achieve this Aim is Distribution, which he argues is justified by retributivism.<sup>130</sup> Hart argues that it is “perfectly consistent” to apply consequentialist principles to justify the General Aim and retributive principles to justify the Distribution, and that by accepting his theory, punishment theorists in retributivist and consequentialist camps could cease their centuries-old “shadow-fighting.”<sup>131</sup> For proportionality, Hart combines elements from retributivism and consequentialism to establish a floor and a ceiling, or a minimum and a maximum amount of punishment. The consequentialist goal of deterrence establishes the floor.<sup>132</sup> Retributivism’s degree-based negative duty establishes the ceiling.<sup>133</sup> Thus, under Hart’s theory, the appropriate degree of punishment to impose upon an offender must be sufficient to achieve deterrence but must not exceed what the offender deserves. Critics accuse Hart’s theory of “being disproportionately lenient,”<sup>134</sup> because it excludes positive retributivism from its

---

123. See HART, *supra* note 65.

124. See Christopher, *Deterring*, *supra* note 3, at 867-68 (“Among these mixed theories of punishment, the most influential is H.L.A. Hart’s.”).

125. See HART, *supra* note 65, at 8-13.

126. *Id.* at 8.

127. *Id.* at 11.

128. See Christopher, *Deterring*, *supra* note 3, at 869 (“Hart’s mixed theory [has] its own difficulty in setting the amount of punishment.”).

129. See HART, *supra* note 65, at 9 (“[T]he General Justifying Aim of the practice of punishment is its beneficial consequences[.]”).

130. *Id.* at 9 (“[T]he pursuit of this General Aim should be qualified or restricted out of deference to principles of Distribution which require that punishment should only be of an offender for an offence.”).

131. *Id.* at 9

132. See Christopher, *Deterring*, *supra* note 3, at 869 (“[D]eterrence concerns set the lower limit on how much punishment to impose.”).

133. *Id.* (“The negative retributivism component sets the upper limit . . . on how much punishment to impose.”).

134. *Id.*

proportionality inquiry. For example, if deterrence is not achieved by the punishment of a particular offender for a particular offense, then punishment is impermissible, because the deterrence sets the floor for the amount of punishment.<sup>135</sup>

To summarize Hart's theory, the General Justifying Aim is satisfied by consequentialism, Distribution is justified by retributivism, and proportionality combines deterrence and degree-based negative retributivism to establish a range of acceptable degrees of punishment. With respect to the conjoined twins problem, the General Justifying Aim of punishment is unproblematic. Consequentialist goals can justify a system of punishment that punishes murderers. The conjoined twins problem raises questions of individual assignments of punishment. Therefore, Hart's justifications of Distribution and proportionality are the focal points of the conjoined twins problem because the inquiry is whether (and how much) punishment of *Guilty Twin* can be justified when *Innocent Twin* will also necessarily suffer the punishment.

Applying Hart's theory's Distribution component to the conjoined twins problem presents largely the same issues that other versions of permissive retributivism face in their resolutions. Distribution only permits punishment "of an offender for an offense."<sup>136</sup> *Guilty Twin* is an offender who has committed the immoral and illegal offense of murder. *Innocent Twin* is free from guilt. Punishment of *Guilty Twin* is permissible; punishment of *Innocent Twin* is prohibited. Hart's theory does not obligate punishment of each and every guilty person; it merely permits punishment. As described above, critics note that Hart's theory cannot justify punishment of a guilty offender if deterrence is not achieved by the punishment. Because punishment of *Innocent Twin*—who is not an offender who committed an offense—is prohibited under this theory, and punishment of *Guilty Twin* is merely permissible and not obligatory, Hart would likely forgo punishment of *Guilty Twin*. Forgoing punishment of *Guilty Twin* raises the same issues that negative retributivists face.

If proponents of Hart's theory instead elect to punish *Guilty Twin*, as an offender who would be punished for an egregious offense, they violate their duty not to punish *Innocent Twin*. Even apart from violating the "who may we punish" prong of Distribution, they also face considerable trouble in assigning the appropriate degree of punishment. The charged offense is murder. The minimum punishment to deter murder is presumably high, because murder is an offense that society

---

135. *Id.* (explaining that retributivist critics of Hart's theory deem it "disproportionately lenient in . . . not justifying any punishment at all of guilty wrongdoers where deterrence could not be promoted.").

136. HART, *supra* note 65, at 9.

especially wishes to prevent. Protecting this heightened societal interest in preventing murder justifies a greater degree of punishment. The maximum punishment for *Guilty Twin* is also high, because a murder conviction makes one eligible for punishments as severe as life imprisonment or capital punishment under desert-based rationales. Thus, both the floor and the ceiling are elevated for *Guilty Twin*. However, because *Innocent Twin* deserves no punishment, the ceiling is actually no punishment at all. Interestingly, proponents of Hart's theory could make a persuasive argument that the floor can be set with respect to *Innocent Twin*. Recall that punishment of *Innocent Twin* is unjustified in the "who may be punished" prong and that this argument only succeeds in the proportionality prong; thus, by the time this argument is presented, the Hart theorist has already violated his theory. To successfully argue that the floor can be satisfied with respect to *Innocent Twin*, the Hart theorist must show that punishment could deter future offenses. Certainly, specific deterrence would be achieved with respect to *Guilty* and *Innocent Twin*. General deterrence might also be achieved through the direct punishment of *Guilty Twin* and the vicarious punishment of *Innocent Twin*. Therefore, if the Hart theorist avoids the "who may be punished" prong of the Distribution inquiry, he could make an interesting argument that the floor requires a high punishment. However, the ceiling, with respect to *Innocent Twin*, is no punishment at all. Thus, the maximum punishment with respect to *Innocent Twin* requires no punishment, but the minimum could justify a high degree of punishment. Hart recognizes this possibility and emphasizes the importance of the "who may be punished" prong, arguing that, while vicarious punishment of an offender's family could achieve consequentialist objectives, it is unjustified under the retributive principles.

Thus, *Guilty Twin* may permissibly be punished under Hart's theory. *Innocent Twin* must not be punished under Hart's theory. The proponent of Hart's theory must forgo punishing both twins. This gives a disproportionately lenient punishment with respect to *Guilty Twin*—more precisely, it results in the release of a known murderer. Thus, although this mixed theory attempts to reconcile the differences between the two competing, broad theories of punishment and avoid the criticisms lodged against each, the conjoined twins problem shows that it actually falls subject to the criticisms leveled against both theories. Under its consequentialist components, it can justify the intentional punishment of *Innocent Twin*. Under its retributive components, it cannot justify punishment of *Innocent Twin* and does not require punishment of *Guilty Twin*. Although forgoing punishment is a solution to the conjoined twins problem, it demonstrates that Hart's theory is an insufficient justification

of punishment. Under this theory, a known murderer must be released. The conjoined twins problem provides an example of why Hart's theory is disproportionately lenient. However, the conjoined twins problem also exposes an additional vulnerability of Hart's theory. Note again that this theory's purpose is to combine the favorable aspects of both consequentialism and retributivism. Consequentialism can resolve the conjoined twins problem by considering consequences apart from deterrence, including the bad consequences for society if a known murderer goes unpunished by virtue of a biologically built-in defense. In other words, a consequentialist could make a defensible argument that the interests of society (especially with respect to public safety) in not providing a known murderer with a license to literally get away with murder outweigh the bad consequences of punishing an innocent. By holding onto retributivist principles that preclude a satisfactory resolution for the conjoined twins problem, Hart fails in his ultimate goal to create a bulletproof theory that combines the two broad theories. His theory is subject to the same criticisms leveled against negative retributivism.

Mitchell Berman presents a recent mixed theory of punishment, which he terms "an integrated dualist theory of punishment."<sup>137</sup> Rather than taking Hart's approach and dividing the stages of punishment and questions presented at those stages into categories and assigning retributivist or consequentialist justifications to each category, Berman divides cases into two categories. These categories are "core cases" and "peripheral cases."<sup>138</sup> Core cases are cases that are correctly adjudicated so the person being punished is factually and legally guilty.<sup>139</sup> Because core cases involve punishment of an actual offender for an actual offense, the punishment in these cases can be justified under retributive terms.<sup>140</sup> By contrast, peripheral cases are situations where punishment is imposed upon a person who is not factually guilty (e.g., where there is a mistaken identity) or legally guilty (e.g., the conduct fails to satisfy each element of the criminal offense).<sup>141</sup> Peripheral cases, which "all involve genuine . . . error,"<sup>142</sup> can be justified under consequentialist terms.<sup>143</sup> A third type, which Berman calls "degenerate cases," involves intentional punishment of known innocents.<sup>144</sup> Because these

---

137. Mitchell N. Berman, *Punishment and Justification*, 118 *ETHICS* 258, 259 (2008).

138. *Id.* at 260.

139. *Id.*

140. *Id.* ("[C]ore cases of punishment are justified . . . on retributivist grounds.")

141. *Id.*

142. Berman, *supra* note 137, at 262.

143. *Id.* at 260-61 ([P]eripheral cases of punishment are justified . . . on consequentialist grounds.")

144. *Id.* at 262 (discussing degenerate cases, including cases of scapegoating or prosecutorial

punishments infringe on the rights of the innocent who do not deserve to be punished, these cases are not justifiable.<sup>145</sup>

Applying Berman's theory to the conjoined twins problem requires only a brief discussion. Because his theory separates punishments into separate "cases," his theory operates under the separatist presumption that the person being punished is a separate individual. The following application assumes that the twins are two persons and treats punishment of each as a separate case.<sup>146</sup> Punishment of *Guilty Twin* constitutes a core case, justified under retributivist principles. Because *Guilty Twin* is factually and legally guilty of murder, his punishment is justified. But punishing *Guilty Twin* requires the Berman proponent to open a degenerate case with respect to *Innocent Twin*, because *Innocent Twin* is an identifiable, known innocent. Although punishment of *Guilty Twin* is justified, the punishment of *Innocent Twin* is unjustified. Punishment of *Guilty Twin* requires the state to engage in unjustified scapegoating of *Innocent Twin* for the sole purpose of achieving the conviction and punishment of *Guilty Twin*. Thus, Berman's theory fails to justify punishment of the twins. The proponent of the Berman theory must then forgo punishment of the *Guilty Twin*. This exposes Berman's theory to the same criticisms leveled against other impure versions of retributivism, including excessive leniency by requiring the release of *Guilty Twin*, a known murderer.

#### IV. CONCLUSION

Application of the conjoined twins problem to impure versions of retributivism—deontological, consequentialist, threshold, negative/weak, victim-conscious, and mixed—reveals that there is no version of retributivism that can satisfy the problem of conjoined twins. Even where a theory seems to successfully resolve the problem by providing an answer (usually declining to punish), that resolution exposes the theory to additional criticisms. The conjoined twins problem is either unresolvable under each theory, or it exposes flaws within the theory. No desert-based justification for punishment can satisfactorily resolve the conjoined twins problem.

---

animus).

145. *Id.* at 280 (“[R]ights infringing would be infliction of suffering that is not genuinely supposed to be deserved, which supports the widespread judgment that . . . degenerate cases of punishment are not justified.”).

146. An attempt to apply Berman's theory another way would be nonsensical, because inquiry into the factual and legal guilt of each individual is necessary in a theory that applies different justification theories to different types of cases (those involving accurate punishments, those involving mistaken punishments, and those involving deliberately inaccurate punishments).