

**SELECTIVE INCOMPATIBILISM, FREE WILL, AND THE
(LIMITED) ROLE OF RETRIBUTION IN PUNISHMENT THEORY**

*Luis E. Chiesa**

TABLE OF CONTENTS

I. INTRODUCTION	977
II. COMPETING APPROACHES TO THE FREE WILL PROBLEM	981
A. <i>(Metaphysical) Libertarianism</i>	981
B. <i>Free Will Skepticism</i>	983
C. <i>Compatibilism</i>	986
III. THE BURDEN OF PROOF PROBLEM AND THE ASYMMETRIES BETWEEN INCULPATION AND EXCULPATION	990
A. <i>The Burden of Proof Problem and the Free Will Debate as Dialectical Stalemate</i>	990
B. <i>Asymmetries Between Exculpation and Inculpation</i>	993
IV. A NEW DEFENSE OF AN OLD THEORY: SELECTIVE INCOMPATIBILISM AND MIXED THEORIES OF PUNISHMENT	995
A. <i>The Connection Between Selective Incompatibilism and Mixed Theories of Punishment</i>	995
B. <i>Selective Incompatibilism, “Confidence Levels,” and the Role of Retribution in Punishment Theory</i>	997
V. CONCLUSION	1000

I. INTRODUCTION

In his important *Rethinking Punishment*, Leo Zaibert provides a novel and thought-provoking defense of retribution.¹ Although I am not a retributivist, there is much to learn from Zaibert’s defense of the role of retribution within a pluralistic account of punishment. One of the virtues of Zaibert’s treatment of the topic is that it incorporates arguments from a variety of important theorists, including some that

*Professor of Law and Director of the Buffalo Criminal Law Center. University at Buffalo School of Law, The State University of New York.

1. LEO ZAIBERT, *RETHINKING PUNISHMENT* (2017) [hereinafter *RETHINKING PUNISHMENT*].

have not been traditionally associated with discussions about punishment. In particular, Zaibert's deployment of G.E. Moore's discussion of organic wholes as a vehicle to help us better understand the role that desert-judgments play in everyday life is as fascinating as it is illuminating.²

Given the breadth and depth of *Rethinking Punishment*, I was somewhat surprised to find that there is little discussion of whether humans have the kind of free will that justifies the imposition of retributive punishment. I am only somewhat surprised because discussions of free will are conspicuously absent in most modern discussions of retributive punishment, at least amongst criminal law theorists.³ This is unfortunate, since the challenges posed to retributive theories of punishment by the free will problem are quite formidable. Succinctly, the challenge is as follows: Most scientists and philosophers agree that the universe and all objects contained within it behave in accordance to the thesis of causal determinism (determinism).⁴ Determinism is the thesis that all events that take place are entirely the product of past events in combination with the operation of physical laws.⁵ At first glance, determinism seems in tension with free will, for how can we have free choice if all of our choices are dictated by past events in conjunction with the operation of immutable physical laws? Since we lack control over both past events and natural laws, it is not unreasonable to conclude that if determinism obtains, we would lack control over our choices too. And, crucially, if we lack control over our choices, then it is plausible to believe that we cannot be blamed for them.

2. *Id.* at 27.

3. Moral philosophers often grapple with questions of free will and punishment. *See, e.g.*, DERK PEREBOOM, *LIVING WITHOUT FREE WILL* (2003); JOHN MARTIN FISCHER & MARK RAVIZZA, *RESPONSIBILITY AND CONTROL: A THEORY OF MORAL RESPONSIBILITY* (1998); ROBERT KANE, *THE SIGNIFICANCE OF FREE WILL* (1996) [hereinafter *SIGNIFICANCE OF FREE WILL*]. In contrast, most criminal law theorists seldom engage seriously with this literature. Even more surprisingly, most criminal law theorists who describe themselves as retributivists either do not engage with the topic at all, or engage with it in superficial ways, often by uncritically adopting some philosopher's views on the subject. *See, e.g.*, DOUGLAS HUSAK, *PHILOSOPHY OF CRIMINAL LAW* 78–122 (1987); LARRY ALEXANDER & KIMBERLY KESSLER FERZAN, *CRIME AND CULPABILITY* 15 (2009) (adopting Stephen Morse's views on the subject).

4. *See generally* Carl Hoerfer, *Causal Determinism*, *THE STANFORD ENCYCLOPEDIA OF PHILOSOPHY* (Jan. 21, 2016), <https://plato.stanford.edu/archives/spr2016/entries/determinism-causal>. Causal determinism holds for objects that are larger than atoms. Subatomic particles, in contrast, behave probabilistically under the laws of quantum mechanics rather than deterministically. *Id.* But, as I will explain in Part II.B of this Article, the probabilistic behavior of subatomic particles pursuant to the rules of quantum mechanics present as much of a challenge to retributive theories of punishment.

5. *Id.*

Finally, since retribution presupposes blame, and determinism calls into question whether anyone can ever be blamed for anything, determinism threatens to unravel the theoretical edifice upon which retributive theories of punishment are erected.⁶

To be sure, a punishment retributivist has several plausible arguments that could allow them to fend off the challenges that determinism poses for retributive punishment. What is surprising, though, is not that punishment retributivists such as Zaibert are unfazed by the threat posed by determinism, but rather that they tend to cavalierly ignore the problem that it presents for retributive punishment. Unfortunately, *Rethinking Punishment* continues with this trend. As such, there is no meaningful discussion of the tension between the thesis of causal determinism and retributive theories of punishment in the body of the book.

Nevertheless, to Zaibert's credit, there is a somewhat related discussion in the Appendix of the book. There, Zaibert argues against a view defended by Derek Parfit that Zaibert somewhat deridingly calls "compatibilism a la carte."⁷ In a nutshell, compatibilism a la carte posits that there is an asymmetry between desert judgments about blame and praise and desert judgments about suffering.⁸ More specifically, Parfit claims that people can deserve blame and praise, but that they can never deserve suffering.⁹ This view follows from Parfit's belief that although humans possess enough freedom to be blamed and praised for how they act, they do not have the degree of freedom that is necessary to justify inflicting suffering on others as a consequence of their choices.¹⁰

Zaibert finds this view puzzling, since it suggests that determinism is compatible with the notion of deserved blame and praise but incompatible with the notion of deserved suffering.¹¹ Zaibert then accuses Parfit of selectively and somewhat arbitrarily choosing some things that one can deserve (everything except suffering) and some things that one cannot deserve (suffering).¹² Zaibert thus argues that "Parfit cherry-picks when compatibilism operates and when it does not."¹³ In order for this cherry-picking to be justified, Zaibert argues that Parfit needs to

6. LEO ZAIBERT, PUNISHMENT AND RETRIBUTION 195 (2006). While there are many ways of defining retribution, I follow Zaibert in defining it as the belief that the imposition of deserved punishment is non-instrumentally or intrinsically good. *Id.*

7. RETHINKING PUNISHMENT, *supra* note 1, at 244.

8. *Id.*

9. *Id.*

10. *Id.*

11. *Id.* at 246.

12. *Id.*

13. *Id.*

demonstrate that there is “something special . . . that would show suffering (and suffering alone) to be undeservable.”¹⁴

Given the strange nature of Parfit’s claim that suffering is the only thing in the world that cannot be deserved, Zaibert unsurprisingly argues that one cannot find that which would show suffering and only suffering to be undeserved.¹⁵ Furthermore, he shows, quite convincingly, that Parfit cannot avail himself of the classic free-will skeptic argument that determinism rules out the possibility of desert, because the truth of this proposition would rule out *all desert judgments*, not just judgments about deserved suffering, as Parfit would like to have it.¹⁶

In what follows, I will not attempt to rehabilitate Parfit’s idiosyncratic defense of what Zaibert calls “incompatibilism a la carte”.¹⁷ I agree with Zaibert that Parfit offers little support for his view and that the support that Parfit tries to muster from Kant’s musings on determinism do not do the heavy lifting that needs to be done in order to legitimate his view. Be that as it may, I think that there is a kind of selective approach to compatibilism that is defensible. I will call this novel approach “selective incompatibilism.”

In a nutshell, the view that I will call selective incompatibilism holds that we should invoke an actor’s desert (or lack thereof) as a reason for exculpation but not as a reason for inculpation. In terms of punishment theory, the view leads to a mixed theory of punishment that allows for retributive judgments about desert to serve as *limits* on the amount of punishment that can be imposed but that disallows such judgments as positive justifications for the imposition of a criminal sanction. Expressed more simply, selective incompatibilism argues that determinism ought to be viewed as compatible with desert for the purposes of exculpating actors, but that it ought to be viewed as incompatible with desert for the purposes of inculpating actors. The end result is a theory of punishment that justifies the imposition of punishment on non-desert grounds (such as deterrence or incapacitation) but that sets desert-based constraints on the kinds of actors that can be punished and the maximum penalties that can be imposed.

This Article is comprised of three parts. Part II compares and contrasts three competing approaches to the free will problem. Part III argues that the free will debate has reached a dialectical stalemate for which there is no end in sight. Part IV introduces and defends a mixed

14. *Id.*

15. *Id.* at 246–47.

16. *Id.* at 247.

17. *Id.* at 244.

theory of punishment based on “selective incompatibilism” as a provisional way out of the free will impasse. A brief conclusion follows.

II. COMPETING APPROACHES TO THE FREE WILL PROBLEM

A. (*Metaphysical*) *Libertarianism*

Perhaps the most intuitive approach to the free will problem is metaphysical libertarianism. Metaphysical libertarianism is different from political libertarianism. Political libertarianism is the view that government should play a minimal role in societal life, usually relegated to providing the bare minimum of essential public services, such as national defense and law enforcement.¹⁸ Political libertarianism takes liberty to be the most basic of human rights and looks at any governmental interference with individual freedom with great suspicion.¹⁹ The most sophisticated defense of political libertarianism is still the one developed by Robert Nozick in his much-celebrated *Anarchy, State and Utopia*, in which he made the case for a minimal state tasked primarily with administering courts of law and with protecting the citizenry against fraud, theft and unjustified uses of force.²⁰

In contrast, metaphysical libertarianism is not a political theory.²¹ Instead, it is a theory about the nature of free will and its relationship to determinism.²² In a nutshell, metaphysical libertarians believe that free will is incompatible with determinism.²³ As a result, scholars who advocate for this view are often called “incompatibilists.”²⁴ The view stems from the belief that for a choice to be truly free it must be the case that the person could have chosen otherwise.²⁵ Determinism is thought by metaphysical libertarians to foreclose the capacity to choose otherwise.²⁶ If, as determinism holds, all events that happen in the universe, including human acts, are the product of events that happened in the past in conjunction with the operation of physical laws, then it is difficult to see how things could be otherwise than they actually are.²⁷

18. Bas van der Vossen, *Libertarianism*, STANFORD ENCYCLOPEDIA OF PHILOSOPHY (Jan. 28, 2019), <https://plato.stanford.edu/archives/spr2019/entries/libertarianism/>.

19. *Id.*

20. ROBERT NOZICK, *ANARCHY, STATE AND UTOPIA* (1974).

21. See JOHN MARTIN FISCHER ET AL., *FOUR VIEWS ON FREE WILL* 3 (2007).

22. *Id.*

23. *Id.*

24. *Id.*

25. Laura Waddell Ekstrom, *Libertarianism and Frankfurt-Style Cases*, in *THE OXFORD HANDBOOK OF FREE WILL* 309, 321 (Robert Kane ed., 2002).

26. *Id.*

27. *Id.*

Since humans lack control over both the past and the operation of natural laws, determinism would seem to rule out our capacity to make choices that are different from the choices that we actually make.²⁸

Given that metaphysical libertarians believe that free will and desert are incompatible with determinism, they argue that humans would lack free will if it turns out that the universe actually works in predictably deterministic ways.²⁹ Nevertheless, metaphysical libertarians hold that we do, in fact, have free will because the thesis of causal determinism is false, at least with regard to its applicability to human conduct.³⁰ In order to reach this conclusion, metaphysical libertarians posit that human choices are plausibly the product of indeterministic processes that, by definition, are not fully accounted for by determinism.³¹ Some, like philosopher Robert Kane, speculate that quantum-mechanical processes may account for salient features of the human decision-making process, especially when it comes to complex decisions.³² Given that the underlying rules of quantum mechanics are probabilistic rather than deterministic, this account of human conduct allows for the possibility that at least some choices are not determined by causal forces that humans cannot control.³³ If the physical reality of the processes underlying human decision-making are not fully deterministic, then there is breathing room for the human will to tilt the balance in favor of a certain choice over another.³⁴ This, argues Kane, could allow for the existence of at least some actions that can be freely willed.³⁵ Others, like legal scholar Ken Levy, argue for a dualistic account of human conduct, in which the physical body, including the brain, is fully subjected to deterministic forces, while the mind is identified as a non-physical entity that is not subject to the constraints of determinism.³⁶ If, as Levy speculates, free choices originate in the operations of the mind and if, as Levy suggests, the mind is a non-corporeal entity that is not constrained by causal determinism, then it is plausible to argue that at least a subset of human behavior can be described as freely willed.³⁷

28. *Id.*

29. *Id.*

30. Robert Kane, *Libertarianism*, in *FOUR VIEWS ON FREE WILL* 5, 7 (2007).

31. *Id.* at 26–28.

32. *Id.* at 29–30.

33. *Id.*

34. *Id.*

35. *Id.* at 30–31.

36. *See generally* KEN LEVY, *FREE WILL, RESPONSIBILITY, AND CRIME: A PHILOSOPHICAL INTRODUCTION* (2019).

37. *Id.*

While we do not currently have sufficient information regarding either the ultimate nature of the human mind³⁸ or the laws that govern the universe at a sub-atomic level, metaphysical libertarian theories such as the ones espoused by Kane and Levy cannot be definitively dismissed. However, it is also fair to say that there is little scientific support for these theories. Although we cannot rule out that, as Kane suggests, the mind is a physical entity that at least partially escapes the grip of determinism, or that—as Levy claims—the mind is a non-corporeal entity that stands fully outside determinism, no current scientific account of the mind suggests that either of these theories is likely to be true. For these reasons, most scholars who share the metaphysical libertarian belief that determinism is incompatible with free will end up rejecting metaphysical libertarianism and instead embrace the view that I will now turn to, which I will describe as “free-will skepticism.”

B. Free Will Skepticism

The scholars that defend the view that I will call “free will skepticism” agree with metaphysical libertarians that determinism seems to rule out moral responsibility.³⁹ Therefore, like metaphysical libertarians, free will skeptics are incompatibilists.⁴⁰ However, unlike metaphysical libertarians, free will skeptics believe that the thesis of causal determinism is likely to be true.⁴¹ Since free will skeptics believe that it is likely that determinism obtains, they contend, as their label suggests, that we have good reasons to be skeptical about whether humans have free will.⁴²

The most robust defense of free will skepticism in contemporary debates is that of philosopher Derk Pereboom. According to Pereboom,

38. It is somewhat embarrassing to admit that in spite of the considerable advances in natural sciences during the last hundred years, we are not even close to understanding how the mind works at the most fundamental level. In fact, we know very little about where consciousness comes from or why consciousness even evolved in the first place. So, while most philosophers and scientists are skeptical of dualist theories of mind such as the one posited by Levy—a skepticism that I share—we are not yet in a position to conclusively rule out such theories.

39. Free will skeptics are often called “hard determinists” or, less often, “hard incompatibilists.” I follow an increasing trend that refers to this view as “free will skepticism” because it is more descriptive than “hard determinism” or “hard incompatibilism.” For examples of use of “free will skepticism” to describe the position that in the past was called hard determinism, see *FREE WILL SKEPTICISM IN LAW AND SOCIETY: CHALLENGING RETRIBUTIVE JUSTICE* (Elizabeth Shaw, Derk Pereboom, & Gregg D. Caruso eds., 2019).

40. PEREBOOM, *supra* note 3, at xiv.

41. *Id.* at 148.

42. *Id.* at 39.

the truth of determinism would rule out what he calls “basic desert.”⁴³ Basic desert is often associated with the degree of control necessary in order to establish robust judgments of deserved blame (and praise),⁴⁴ such as the judgments of deserved blame that form the backbone of retributive theories of punishment.⁴⁵ While some free will skeptics reject the idea of moral responsibility as incoherent,⁴⁶ some argue that moral responsibility is a coherent idea, but it is unsupported by prevailing scientific theories.⁴⁷ Regardless of this internal disagreement, all free will skeptics agree that current responsibility practices that depend on judgments of blame and/or praise are unjustified in light of either the truth of the thesis of causal determinism, or in light of current scientific understandings of our conscious decision-making processes.⁴⁸

Free will skeptics believe that desert judgments require that humans exercise a degree of control over their thoughts or actions.⁴⁹ They assume, correctly I believe, that we cannot be fairly blamed for acts or thoughts that we cannot sufficiently control.⁵⁰ This would explain why we do not tend to blame small children in the basic desert sense, since we believe that infants are not yet mature enough to control their behavior in a way that would justify robust judgments of blame.⁵¹ It would also explain why we do not blame animals for their behavior in the same way that we blame humans. The underlying assumption seems to be that animals lack the kind of control that is necessary for us to hold them morally responsible for their conduct. Free will skeptics are happy to concede that most of us believe that humans, unlike infants or animals, do possess the degree of control necessary to ground judgments of blame in the basic desert sense.⁵² They argue, however, that this belief, while widely shared, is ultimately unjustified because it is in tension with what we know about how the universe works.⁵³

As far as we know, the universe and all the things contained within it behave in accordance with deterministic rules, at least at the macroscopic level.⁵⁴ As such, the behavior of objects is fully determined

43. *Id.* at 160.

44. *Id.* at xx–xxi.

45. *Id.* at 160.

46. *Id.* at xviii.

47. *Id.* at 197.

48. *See, e.g.*, SIGNIFICANCE OF FREE WILL, *supra* note 3 at 230; PEREBOOM, *supra* note 3, at xix; Hoerfer, *supra* note 4, at 3.

49. PEREBOOM, *supra* note 3 at xxi.

50. *Id.* at xvi.

51. *See, e.g.*, SIGNIFICANCE OF FREE WILL, *supra* note 3 at 69.

52. *See* PEREBOOM, *supra* note 3, at 86–88.

53. *See id.* at 87.

54. *See id.* at 71.

by a combination of past events and the operation of physical laws.⁵⁵ Since such objects do not have the ability to change the past or the operation of physical laws, it makes little sense to hold these objects morally responsible for the consequences of their behavior.⁵⁶ Nevertheless, humans have historically believed themselves to be exceptions to this otherwise universal rule of behavior.⁵⁷ According to the conventional account, humans possess a special kind of feature, ability or power that allows them to escape from the grip of determinism.⁵⁸ But, as free will skeptics argue, there is no evidence that humans are, in fact, an exception to the rule of causal determinism.⁵⁹

Some contemporary metaphysical libertarians respond to this objection by positing that human choices may be the product of indeterministic processes such as the ones that seem to explain the behavior of subatomic particles.⁶⁰ Perhaps the neuronal firings that generate thoughts and culminate in the conscious choices that will later form the basis for judgments about moral responsibility respond to indeterministic quantum mechanical processes rather than to deterministic physical laws.⁶¹ In response, free will skeptics once more argue that there is no scientific basis to conclude that this is how the mind works.⁶² But beyond this reply, free will skeptics argue that there is good reason to be skeptical about moral responsibility even if it turns out that the workings of the mind are the product of indeterministic processes as metaphysical libertarians such as the ones that Kane suggests.⁶³ Indeterminism cannot serve as the basis for moral responsibility because indeterminism ultimately boils down to randomness, and randomness threatens moral responsibility as much, or perhaps more, than determinism.⁶⁴ Just like we cannot claim responsibility for conduct that is the product of deterministic causal

55. *See id.*

56. *See id.*

57. *See id.*

58. *See id.* at 69–70.

59. *See id.* at 71–73.

60. *See, e.g., id.* at 71–73.

61. *See generally* SIGNIFICANCE OF FREE WILL, *supra* note 3, at 2 (noting worthiness of discussing compatibility of free will and determinism, as “it could turn out that future physicists conclude that the universe is deterministic, contrary to the contemporary consensus about at least quantum mechanics”).

62. *See, e.g.,* PEREBOOM, *supra* note 3, at 69 (“Some libertarians maintain that there could be no natural scientific account of agent-causal powers.”).

63. *Id.* at 71.

64. *Id.* at 71–72.

factors that we cannot control, so too can we not claim responsibility for conduct that results as a consequence of pure luck or randomness.⁶⁵

While the arguments deployed by free will skeptics in support of their position are plausible enough, many argue that we should not be too quick to abandon the notion of moral responsibility.⁶⁶ Two main arguments have been presented against free will skepticism. The first points out that the foundation of judgments of deserved blame and praise is *not* the capacity to do otherwise or the capacity to control behavior to a certain degree.⁶⁷ Instead, some critics of free will skepticism argue that the basis of desert is to be found in some other feature of human conduct, such as the capacity for rationality.⁶⁸ Other critics suggest that adopting the free will skeptic view would lead us to reject many features of societal life that are essential to human flourishing, such as the social practice of holding people accountable for their wrongdoing, and the interpersonal practices of expressing love, resentment, blame, and praise.⁶⁹ Since the free will skeptic view would lead to abandoning practices and attitudes that are essential to human flourishing, we should not embrace this view unless we are definitely convinced that no other view is plausible. But we are not even close to being convinced that free will skepticism's view of the world and human behavior is accurate. While the free will skeptic's view is plausible, it is by no means definitively and uncontroversially true. If so, we should continue to hold on to our practices of blaming, praising, and holding people accountable for their wrongdoing until we become convinced that there is no justification for these practices.⁷⁰

C. Compatibilism

Compatibilism is the view that moral responsibility is not in tension with the thesis of causal determinism.⁷¹ Therefore, compatibilists, unlike

65. See *id.* at 72 (“If everything is wholly constituted of microphysical entities governed by [deterministic or fundamentally statistic laws of physics], no room is left for agent-caused free decisions.”).

66. See, e.g., JOHN MARTIN FISCHER & MARK RAVIZZA, S.J., *RESPONSIBILITY AND CONTROL* (1998).

67. PEREBOOM, *supra* note 3, at 106–07.

68. See, e.g., Stephen J. Morse, *Reason, Results, and Criminal Responsibility*, 2004 U. ILL. L. REV. 363, 375 (2004).

69. See, e.g., PEREBOOM, *supra* note 3, at 206.

70. A prominent free will skeptic reply to this objection is that free will skepticism does not require abandoning the practice of holding each other to account for wrongdoing nor does it require jettisoning certain attitudes, such as love and admiration. See *generally id.* at 198–215.

71. See Michael McKenna & D. Justin Coates, *Compatibilism*, STANFORD ENCYCLOPEDIA OF PHILOSOPHY (Feb. 25, 2015) <https://plato.stanford.edu/archives/win2018/entries/compatibilism/>.

metaphysical libertarians and free will skeptics, argue that determinism is compatible with desert and free will.⁷² Furthermore, compatibilists argue that determinism poses no threat to moral responsibility. While compatibilists agree with the free will skeptic critique of metaphysical libertarianism as unscientific, they disagree with the free will skeptic view that there would be no justification for basic desert judgments of blame and praise if causal determinism obtains.⁷³ Most philosophers and legal scholars who tackle the free will problem advocate for some kind of compatibilist account of the relationship between determinism and moral responsibility.⁷⁴

Many different varieties of compatibilism find support in the contemporary literature on free will. In spite of the variances amongst different compatibilist theories, most attempt to identify one or several distinctively human traits that can serve as the basis for desert-based judgments about moral responsibility regardless of whether determinism obtains or not.⁷⁵ As I alluded to in the previous subsection, a particularly popular approach amongst at least a subset of compatibilists is to argue that judgments of deserved blame and praise are justified if they are directed at a being that has capacity for rationality regardless of whether the being had the capacity to act or choose otherwise.⁷⁶ The underlying idea is that morality is a system of norms that provide its addressees with reasons in favor or reasons against acting in a certain kind of way.⁷⁷ For example, the moral rule against causing the death of non-aggressors can be thought of as a norm that seeks to provide us with reasons against killing innocent people. Once moral norms are thought of as reasons that are designed to guide our conduct, then we can think of the capacity to grasp and be guided by reasons as the capacity that allows for judgments of blame or praise for breaking or following moral norms.⁷⁸ This is the view defended by legal scholar Stephen Morse, who contends that we can be held morally responsible for violating moral norms if we have the general capacity to grasp reasons for action and to be guided by such reasons for action.⁷⁹ That is, we are justified in holding people morally responsible when they have the general capacity to grasp and be guided by reasons but nevertheless fail to modify their behavior so as to act in a

72. See, e.g., PEREBOOM, *supra* note 3, at xiv.

73. *Id.* at 139–57.

74. *Id.* at xiv.

75. *Id.*

76. See, e.g., Morse, *supra* note 68, at 441.

77. See *id.* at 368.

78. See *id.* at 368.

79. *Id.*

way that is compatible with the reasons given by the norm for acting in certain ways.⁸⁰

Instead of attempting to identify features of human conduct that can serve as the basis for moral responsibility in a deterministic world, some compatibilists prefer to focus on asking whether it is even possible to lead meaningful lives without ever assigning moral blame and responsibility.⁸¹ For some, like philosopher P.F. Strawson, certain kinds of attitudes and emotions are essential to the human condition, and we could not live without them even if we tried.⁸² Strawson called these emotions the “reactive attitudes.”⁸³ They include, but are not limited to, feelings of resentment, gratitude, forgiveness, anger, and love.⁸⁴ These attitudes are essential to the creation and maintenance of meaningful interpersonal relationships.⁸⁵ As such, abandoning them would lead not only to an attitudinal void, but also to a considerable impoverishment of the kinds of interpersonal relationships that make societal life worthwhile.⁸⁶ Scholars such as Strawson argue that we hold people morally responsible not because they are truly morally responsible, as there is no such thing as “true” moral responsibility.⁸⁷ Instead, we hold people responsible because doing so is necessary for constituting meaningful relationships with others.⁸⁸ Ascriptions of moral responsibility would thus appear as a useful social practice that is not in need of external justification. Pursuant to this view, the truth or falsity of the thesis of causal determinism ought to have little bearing on our moral responsibility practices.

Furthermore, and perhaps more importantly, our actual practices of ascribing moral responsibility do not seem to be particularly responsive to determinism. For example, our current practices of blaming and punishing do not provide an excuse for “lack of control.” In fact, a perfunctory survey of American criminal law seems to suggest the opposite.⁸⁹ Courts routinely reject excusing defendants on grounds of severe social disadvantage or deprivation, even when faced with evidence

80. *See id.* at 441.

81. *See id.* at 443.

82. Peter Strawson, *Freedom and Resentment*, in *ETHICAL THEORY: AN ANTHOLOGY* 340, 342–43 (Russ Shafer-Landau ed., 2d ed., 2013).

83. *Id.* at 343.

84. *See id.* at 342, 344.

85. *See id.* at 344.

86. *See id.*

87. *See, e.g., id.* at 349.

88. Some legal theorists, such as Michael Moore, have flirted with similar kinds of arguments. *See, e.g.,* Michael S. Moore, *Causation and the Excuses*, 73 CALIF. L. REV. 1091, 1139–40 (1985).

89. *See, e.g., id.* at 1092.

that severe social deprivation is quite criminogenic, and, therefore, highly correlated with crime.⁹⁰ Brainwashing defenses have suffered a similar fate.⁹¹ Even in the context of mental illness, the contemporary trend is to reject volitional or lack of control tests for incapacity such as the so-called irresistible impulse test.⁹² Compatibilists like Stephen Morse point to this feature of law to argue that lack of control has never been essential to judgments about desert, blame and moral responsibility.⁹³ Instead, he argues that what the law truly cares about is cognitive abilities.⁹⁴ This is perhaps most obvious in the context of insanity, where cognitive impairments that prevent us from understanding moral or legal norms as reasons for action preclude blame and punishment, whereas volitional impairments generally do not.⁹⁵

Metaphysical libertarians and free will skeptics argue that the kind of desert-based blame judgments that compatibilists argue are compatible with determinism are not the type of desert-based judgments that are necessary to justify moral responsibility.⁹⁶ With regard to the Strawsonian claim that reactive attitudes, such as blame and praise, are essential to having meaningful human relationships and can thus not be abandoned,⁹⁷ libertarians and skeptics point out that the fact that having certain beliefs is necessary for living a fulfilling life does not make the beliefs any more true or accurate.⁹⁸ For example, many people consciously or unconsciously turn to belief in the afterlife as a way of coping with existential or death anxiety.⁹⁹ Even if we became convinced that believing in the afterlife is good for us, it would be a non-sequitur to conclude that such a belief is “true” or “accurate” solely because it is good for us. By the same token, one can believe that ascribing blame is essential for human relationships without being convinced that ascribing blame is truly “just” or “fair.” Another example may be useful here.

90. See, e.g., *United States v. Alexander*, 471 F.2d 923, 959 (D.C. Cir. 1972).

91. See Richard Delgado, “*Rotten Social Background*”: *Should the Criminal Law Recognize a Defense of Severe Environmental Deprivation?*, 3 *LAW AND INEQ.* 9, 19 (1985).

92. See, e.g., *United States v. Lyons*, 731 F.2d 243, 247–48 (5th Cir. 1984) (jettisoning the volitional test for insanity).

93. See, e.g., Stephen J. Morse, *Criminal Responsibility and the Disappearing Person*, 28 *CARDOZO L. REV.* 2545, 2552–53 (2007).

94. *Id.* at 2553–54.

95. See *id.* at 2551–53.

96. See *id.* at 2547–48.

97. Strawson, *supra* note 82, at 342–43, 346.

98. See, e.g., Larissa Rainey, *The Search for Purpose in Life: An Exploration of Purpose, the Search Process, and Purpose Anxiety* (Aug. 1, 2014) (unpublished masters capstone project, University of Pennsylvania) (on file with University of Pennsylvania Library).

99. See *id.*

Perhaps levying some kind of tax on the population is necessary for a well-functioning society, but that does not show that imposing such taxes is “just.” In short, one has to be careful not to confuse utility with justice. Doling out retributive punishment may be a useful way of fostering social cohesion by channeling the power of our reactive attitudes,¹⁰⁰ but that alone does not make imposition of retributive punishment just.

III. THE BURDEN OF PROOF PROBLEM AND THE ASYMMETRIES BETWEEN INculpATION AND EXculpATION

A. *The Burden of Proof Problem and the Free Will Debate as Dialectical Stalemate*

It is all too common in debates about free will for participants to reach an impasse. An obvious reason for this is substantive. Compatibilists argue that they have identified human capacities—such as the capacity for rationality—that can justify ascriptions of moral blame even if determinism obtains.¹⁰¹ Incompatibilists disagree,¹⁰² contending that true ascriptions of blame and responsibility are justified only if we have the capacity to choose otherwise.¹⁰³ They further argue that the truth of determinism would rule out the existence of such a capacity.¹⁰⁴ A dispassionate observer assessing the arguments in favor of compatibilism and incompatibilism will probably come away feeling that both positions are plausible and that advocates of each view score points against the other. It thus seems that we are at a point when the only thing that we can say for sure about the free will debate is that there is reasonable disagreement amongst members of different camps and that there is no obvious way out of the disagreement. The substantive debate

100. This is, in fact, the way in which Paul Robinson argues we should use desert in the context of punishment theory. See generally Paul H. Robinson, *Empirical Desert*, in CRIMINAL LAW CONVERSATIONS 29 (Paul H. Robinson et al. eds., 2009) (suggesting, pursuant to the view Robinson calls “empirical desert,” that we should punish according to desert not because doing so comports with some philosophical standard of justice, but rather because punishing in accordance to desert maximizes utility).

101. Morse, *supra* note 93, at 2551.

102. As I explained in the previous subsection, incompatibilists can be subdivided into metaphysical libertarians that believe that determinism does not fully explain human choices and behavior, and free will skeptics who believe that human action and choices are fully determined by factors that lie beyond our control. See Kane, *supra* note 30, at 22–23; PEREBOOM, *supra* note 3, at xiii.

103. Morse, *supra* note 93, at 2548.

104. *Id.*

about free will has thus reached a “dialectical stalemate” for which there seems to be no resolution in sight.¹⁰⁵

The substantive disagreement between compatibilists and incompatibilists is well-documented.¹⁰⁶ There is, however, an additional source of disagreement that is best described as procedural rather than substantive. Unlike the substantive disagreements between compatibilists and incompatibilists, the procedural disagreement between them remains undertheorized. The procedural disagreement takes place primarily between compatibilists and free will skeptics and it chiefly relates to which of these two camps has the burden of proof with regard to their claims.

Compatibilists tend to argue that free will skeptics have the burden of proving that their views are true for at least three reasons.¹⁰⁷ First, compatibilists argue that free will skeptics have the burden of proof because their views are counterintuitive.¹⁰⁸ They claim that free will skeptic views run counter to intuition because the vast majority of people seem to assume that we can freely make certain choices and that we should be blamed or praised for the choices we make.¹⁰⁹ Second, compatibilists argue that free will skeptics should have the burden of proof because their position runs counter to the assumptions of moral and legal systems throughout history.¹¹⁰ As I explained in the previous subsection, legal systems have historically assumed that humans have freedom of choice and that they can be blamed for their wrongful choices.¹¹¹ The same seems to be true of many moral systems at different historical times and in a multitude of places.¹¹² Finally, compatibilists

105. A dialectical stalemate arises when proponents of an argument cannot muster sufficient evidence to decisively prove the argument, while the argument’s detractors cannot gather sufficient evidence to decisively refute it. JOHN MARTIN FISCHER, *OUR STORIES: ESSAYS ON LIFE, DEATH, AND FREE WILL* 116–17 (2009). In the free will context, those who claim that humans have enough control over their actions to be considered morally responsible for their conduct have failed to conclusively demonstrate either that determinism is probably false or that the truth of determinism does not undermine their claim. John M. Fischer, *Compatibilism*, in *FOUR VIEWS ON FREE WILL* 44, 54 (2007). In contrast, those who suggest that determinism is probably true and that it is incompatible with the free will necessary for moral responsibility have failed to decisively demonstrate the latter. Manuel Vargas, *Revisionism*, in *FOUR VIEWS ON FREE WILL* 126, 151 (2007).

106. Morse, *supra* note 93, at 2547–53 (discussing the disagreement between compatibilists and incompatibilists).

107. Eddy Nahmias et al., *Is Incompatibilism Intuitive?*, 73 *PHIL. AND PHENOMENOLOGICAL RES.* 28, 30 no. 1 (2006).

108. *Id.*

109. *Id.*

110. *Id.* at 31.

111. See generally Morse, *supra* note 93, at 2548.

112. See, e.g., *id.* at 2547.

contend that free will skeptics have the burden of proof because their views tend to falsify much of the attitudes and emotions that characterize our moral lives.¹¹³ As P.F. Strawson forcefully pointed out, the free will skeptic view would lead to abandoning many reactive attitudes that seem to be essential to forging meaningful interpersonal relationships.¹¹⁴ We should not jettison these attitudes unless we are confronted with near incontrovertible proof that expressing these attitudes is morally unjustified. Compatibilists contend that free will skeptics have failed to meet this high burden.¹¹⁵ If so, we are still justified in assuming that ascribing moral responsibility and blame is justified.

In contrast, free will skeptics often argue that the burden of proof should be on compatibilists instead.¹¹⁶ This view is most frequently espoused in discussions about whether state-sanctioned retributive punishment is justified.¹¹⁷ Imposing retributive punishment is legitimate only if we are justified in making desert-based judgments. The problem is that, as both metaphysical libertarians and free will skeptics argue, determinism threatens to undermine the notion of deserved suffering, which lies at the core of retributive theories of punishment.¹¹⁸ Free will skeptics maintain that compatibilists have the burden of proof when they intend to deploy desert-based judgments of blame as a justification for deliberately inflicting suffering on wrongdoers.¹¹⁹ That is, compatibilists should have the burden of proving beyond the shadow of a doubt that desert-based judgments are justified before invoking these kinds of judgments as reasons for punishment.¹²⁰ Since punishment is the worst thing that a democratic state can do to its fellow citizens, free will skeptics argue that we should not impose such a draconian sanction unless we are fully convinced that determinism does not undermine the legitimacy of desert-based judgments.¹²¹ Since reasonable people disagree regarding whether determinism rules out basic desert, compatibilists cannot currently meet their burden of proof, with the consequence that we should abstain from engaging in retributive punishment practices.

113. Nahmias et al., *supra* note 107, at 32.

114. Strawson, *supra* note 82, at 346.

115. See Nahmias et al., *supra* note 107, at 32–33.

116. Vargas, *supra* note 105, at 133; Nahmias et al., *supra* note 107, at 39.

117. Nahmias et al., *supra* note 107, at 35–39.

118. Derk Pereboom, *Hard Incompatibilism*, in *FOUR VIEWS ON FREE WILL* 85, 115 (2007).

119. Nahmias et al., *supra* note 107, at 29.

120. *Id.*

121. *Id.* at 29–30.

It seems to me that, once more, compatibilists and free will skeptics score points against each other when it comes to the question of who has the burden of proof. Compatibilists are right that the free will skeptic position is counterintuitive, runs counter to our history of ascribing blame, and falsifies significant parts of our moral lives. These arguments provide us with good reasons to uphold retributive practices of blaming and praising unless free will skeptics can conclusively demonstrate that such practices are unjustified. Free will skeptics do not appear to have demonstrated this. But free will skeptics also seem to be right when they argue that we should abstain from punishing people solely on the basis that they deserve to suffer unless we have definitive proof that deserved suffering is justified. Given the current state of the free will debate, it appears that we do not have conclusive arguments showing that deserved suffering is, in fact, justified.

It seems, then, that the debate regarding whether compatibilists or free will skeptics have the burden of proving their claims has reached the point of dialectical stalemate in much the same way as the substantive debate between these two camps has produced what appears to be an irreconcilable set of disagreements. How are we to proceed in light of these substantive and procedural impasses? In the next and final section of this Article, I argue for a way out of the stalemate that reflects a compromise between compatibilist and free will skeptic views.

B. Asymmetries Between Exculpation and Inculpation

Instead of categorically arguing in favor or against retributive practices of blaming and praising, in what follows I will argue in favor of a selective use of blame and praise in a way that I hope strikes a reasonable balance between the concerns that animate compatibilists and free will skeptics. This view, which I call “selective incompatibilism,” focuses almost exclusively on the role that desert judgments ought to play in our practices of blaming and punishing. I believe that this focus is justified because punishment theory is where the rubber meets the road in debates about free will. It is one thing to invoke desert as a reason for holding each other accountable in the context of interpersonal relationships, but it is quite a different thing to invoke desert as a reason for intentionally depriving someone of their property, freedom, or even life. While we can more easily indulge our tendency to ascribe blame in the context of our personal relationships, we should be considerably more careful in indulging these inclinations when it comes to assigning blame in the context of state-imposed retributive punishment.

The view I will defend here trades on the asymmetry between inculpatory and exculpatory judgments. In the context of punishment,

inculpatory judgments give us reasons for punishing a person. Exculpatory judgments, on the other hand, give us reasons for abstaining from punishing a person. Our system of criminal justice attaches important consequences to the distinction between inculcation and exculpation. For our purposes, the most important one relates to the burdens of proof. Pursuant to our Supreme Court's reasonable doubt jurisprudence, the prosecution bears the burden of proving the inculpatory elements of criminal conduct beyond a reasonable doubt.¹²² Expressed differently, the prosecution has the burden of proving the elements of a criminal offense beyond a reasonable doubt.¹²³ In contrast, the Supreme Court has held on multiple occasions that the Constitution does not require that the prosecution disprove exculpatory defenses beyond a reasonable doubt.¹²⁴ Instead, the burden of proving exculpation can be lawfully placed on the defendant.¹²⁵ In practical terms, this means that states may constitutionally shift the burden of proving justification and excuse defenses to the defendant.¹²⁶

But why do we require the prosecution to prove the inculpatory facet of the offense beyond a reasonable doubt while we simultaneously allow states to shift to the defendant the burden of proving exculpation? We do this because of the belief that we perpetrate a greater injustice when we wrongfully punish an innocent person than when we wrongfully acquit a guilty person. This is the idea underlying Blackstone's oft-cited assertion that "it is better that ten guilty persons escape than that one innocent suffer."¹²⁷ This intuition is the basis for what I call the inculcation/exculpation asymmetry. Given the severity of the criminal sanction and the value that we place on human freedom, we worry about mistakenly inculcating more than we worry about mistakenly exculpating.

I submit that the same worry that leads us to discriminate between the inculpatory and exculpatory dimensions of criminal conduct for the purposes of burdens of proof ought to lead us to distinguish between these two dimensions for the purposes of punishment theory. Since wrongfully punishing someone is generally believed to be a greater evil than wrongfully acquitting someone, I argue that we ought to be more careful

122. *In re Winship*, 397 U.S. 358, 364 (1970).

123. *Id.*

124. *See Martin v. Ohio*, 480 U.S. 228, 230–31 (1987) (holding that it is in accord with the constitution to impose on the defendant the burden of proving self-defense by a preponderance of the evidence); *Leland v. Oregon*, 343 U.S. 790, 798–99 (1952) (holding that it is in accord with the constitution to impose on the defendant the burden of proving insanity beyond a reasonable doubt).

125. *Martin*, 480 U.S. at 230–31.

126. *See id.*; *Leland*, 343 U.S. at 789–99.

127. 2 SIR WILLIAM BLACKSTONE, COMMENTARIES *342, *358 (1753).

when we invoke desert as a source of inculcation than we should be when we invoke (lack of) desert as a reason for exculpation. But how does this tie in to debates about free will and its compatibilism with determinism? This is the question I take up in the next and final part of this Article.

IV. A NEW DEFENSE OF AN OLD THEORY: SELECTIVE INCOMPATIBILISM AND MIXED THEORIES OF PUNISHMENT

As I mentioned in Part III, it is fair to say that the current free will debate has reached a dialectical stalemate. How are we to proceed in light of this dialectical stalemate? Given our background assumption that it is worse to punish an innocent than to fail to punish a guilty person, I propose the following tentative solution as a way out of the stalemate. For the purposes of justifying the infliction of punishment, we ought to err on the side of caution and assume that determinism is both true and incompatible with free will, and, as a consequence, with desert. On the other hand, for the purposes of limiting the imposition of punishment, we ought to err on the side of freedom and assume that determinism is compatible with free will, and, therefore, with desert. I will call this view selective incompatibilism, since it holds that desert is compatible with determinism for the purposes of freedom-enhancing defenses but not for the purposes of freedom-constraining inculcation.

A. The Connection Between Selective Incompatibilism and Mixed Theories of Punishment

In a nutshell, selective incompatibilism leads to what punishment scholars call a “mixed theory of punishment.”¹²⁸ A mixed theory of punishment is an approach that justifies punishment by appealing to a combination of different justifications or purposes.¹²⁹ The most popular kind of mixed theory of punishment is one that combines consequentialist and non-consequentialist justifications for punishment.¹³⁰ A common way of doing so is by arguing that the chief purpose of punishment should be to deter crime, but that the punishment imposed should never exceed the punishment that is deserved.¹³¹ This approach combines a consequentialist theory of punishment—deterrence—with retribution, a non-consequentialist justification for punishment.¹³²

128. See Russell Christopher, *Deterring Retributivism: The Injustice of “Just” Punishment*, 96 NW. L. REV. 843, 867 (2002).

129. See *id.* at 867–69.

130. *Id.*

131. *Id.* at 868–69.

132. *Id.* at 868.

Mixed theories of punishment are often criticized for being arbitrary or internally inconsistent.¹³³ At first glance, there is something ad-hoc about mixing and matching consequentialism with deontological theories of punishment in the way that mixed theories require. Furthermore, retributivists contend that mixed theories are not really mixed, since they incorporate “negative” retributivism but entirely leave out “positive” retributivism.¹³⁴ Negative retributivism posits that desert serves as a side-constraint on the punishment imposed.¹³⁵ Under this approach, desert would cap the punishment that can be imposed.¹³⁶ Positive retributivism, on the other hand, serves not as a limit on the amount of punishment that is to be imposed, but instead as an affirmative reason to punish the offender up to the amount of punishment that is deserved.¹³⁷ A mixed theory such as the one described above embraces the upper limit on punishment imposed by negative retributivism, but rejects the affirmative reasons to punish provided by positive retributivism.¹³⁸

Incorporating negative retributivism into a mixed theory of punishment while simultaneously rejecting positive retributivism may strike some as an unjustifiable exercise in cherry-picking.¹³⁹ The cherry-picking objection gains steam when we come to see that defenders of mixed theories of punishment are typically consequentialists that want to avoid some of the strong objections that can be leveled against consequentialist theories of punishment.¹⁴⁰ The main problem with theories of punishment that focus on crime prevention is that they offer no built-in limit regarding who can be punished and how much they can be punished.¹⁴¹ As such, consequentialist theories could, in principle, authorize punishing the innocent or punishing in excess of what is deserved, as long as doing so would maximally deter crime.¹⁴²

I sometimes ask my students: “What would be the most ruthlessly efficient way of deterring speeding infractions?” Few initially land on the right answer: imposing the death penalty for speeding. Consequentialists about punishment could only object to such punishment by arguing that,

133. See, e.g., 104 WHITLEY R.P. KAUFMAN, HONOR AND REVENGE: A THEORY OF PUNISHMENT 73–92 (Francisco J. LaPorte et al. eds., 2013).

134. Christopher, *supra* note 128, at 868–69.

135. *Id.*

136. *Id.* at 869.

137. *Id.* at 865–66.

138. See *id.* at 869.

139. See *id.*

140. See *id.* at 867–69.

141. See *id.* at 890–91.

142. See *id.* at 891–92.

in spite of initial appearances, sentencing to death those who speed would not end up maximizing welfare. But this consequentialist reply rings hollow. After all, there is a more fundamental reason why punishing speeding with the death penalty is wrong: doing so is unjust, regardless of whether it would produce good consequences or not. Retributivists can easily explain the injustice of sentencing speeding offenders to death. Killing those who speed is unjust because speeding simply does not deserve such an extreme form of punishment. In contrast, consequentialists cannot avail themselves of this type of argument. At least in principle, if the consequentialist benefits of punishing speeding with the death penalty outweigh its costs, consequentialists should be compelled to impose such punishment.

In order to avoid such an absurd result, many consequentialists advocate for a mixed theory that incorporates negative retributivism as a way of softening the edges of consequentialist punishment theory in much the same way as capitalists embrace welfare reforms in order to make capitalism more palatable to the masses. So framed, the consequentialist turn to a mixed theory of punishment reeks of desperation. As a result, rather than looking like the product of reasoned deliberation about the aims of punishment, mixed theories often appear as cop-outs by consequentialist philosophers. It is in this context that the cherry-picking objection against mixed theories of punishment rings true.

B. Selective Incompatibilism, “Confidence Levels,” and the Role of Retribution in Punishment Theory

Having built up the cherry-picking objection that can be lodged against mixed theories of punishment, I will now provide a defense of a mixed theory that I believe avoids the thrust of this objection. I assume that most would agree that the minimum level of confidence in a belief that one should have in order to act based on that belief depends on what kind of action one is purporting to take on the basis of said belief.¹⁴³ I assume that most would also agree that the more harmful the act that one intends to take on the basis of a belief, the higher the level of confidence that one should have in the belief. In contrast, the less harmful the act that one is planning to engage in on the basis of a certain

143. My use of “degrees of confidence” or “confidence levels” is meant to convey a similar idea to the term “credence level” in the context of statistical analysis. A credence level is a measure of the strength of a belief expressed in terms of a percentage. Antonio Vallecillo, *Belief Uncertainty in Software Models*, MODELING LANGUAGES (Mar. 12, 2019), <https://modeling-languages.com/belief-uncertainty-models-uml/>. The higher your level of credence, the more confident you are that a certain belief is true. *See id.*

belief, the level of confidence that one should have in the belief can be lower.

Imagine that you are contemplating taking your child, Maria, to Disney as a reward for behaving well while you were away on vacation. Further imagine that you are contemplating not taking your other child, Luis, to Disney for behaving badly while you were away on vacation. Imagine as well that you are contemplating physically punishing your other child, Angel, for behaving really badly while you were on vacation. Finally, imagine that you are not one hundred percent sure that Maria behaved well, that Luis behaved badly, or that Angel behaved really badly. While the babysitter reports that they behaved in the way outlined above, Luis and Angel dispute the babysitter's account of bad behavior. You also have reason to believe that the babysitter could be exaggerating the extent of the bad behavior of Luis and Angel in order to charge more next time that they are asked to babysit the children.

Faced with uncertainty about what happened, what level of confidence in the good or bad behavior of your children while you were on vacation should you have before you reward or punish them? Intuitively, it seems that you should have a higher confidence level in order to justify physically punishing Angel than the confidence level that you should have in order to justify depriving Luis of a trip to Disney. In contrast, it would seem that you are justified in having an even lower confidence level in order to justify rewarding your child with a trip to Disney. If you share these intuitions, then you perceive a sliding scale of confidence levels that gets ratcheted up as we seek to do more harmful things in the name of a certain belief and gets ratcheted down as we seek to do less harmful things or even positive things in the name of the belief.

Selective incompatibilism approaches the role of blame in the context of punishment theory in the same way that a parent approaches confidence levels in the context of rewarding or punishing their children. The confidence level that we ought to have in the belief that people can actually deserve things should be higher when we are invoking desert as a reason for intentionally causing suffering than we invoke it as a reason for abstaining from causing suffering. For this reason, *one should have an extremely high degree of confidence in beliefs that serve as the basis for punishment decisions*. This is one way of understanding the different quantum of evidence that American criminal law requires for establishing that the defendant engaged in an offense versus the quantum of evidence it requires for establishing that the defendant acted in self-defense or in a state of insanity. By requiring that the state prove the commission of an offense beyond a reasonable doubt while simultaneously allowing the defendant to secure an acquittal by proving defenses by a preponderance of the evidence, we are conveying there is

an asymmetry between the confidence levels required to establish inculcation and the confidence levels required to establish exculpation. If we seek to make someone criminally liable, the law authorizes such a finding only when there is an extremely high degree of confidence in culpability.¹⁴⁴ In contrast, if we seek to relieve criminal liability, the law allows us to do so on the basis of a much lower degree of confidence than that which is required to authorize a criminal conviction.¹⁴⁵ Expressed numerically, we can say that the factfinder should have a degree of confidence in the defendant's culpability that exceeds 90% in order for a finding of criminal liability to be appropriate. In contrast, we can assert that a factfinder should absolve the defendant of liability if her degree of confidence that the defendant acted in self-defense exceeds 50%.

The same reasons that lead to an asymmetry in confidence levels in the context of the beyond a reasonable doubt jurisprudence should lead to an asymmetry in the confidence levels that we should have in the context of determining the role that retribution ought to play in punishment theory. If we require an extremely high level of confidence in the belief that a defendant culpably engaged in wrongdoing before we make a finding of criminal liability, we should require a similar degree of confidence in the justifiability of our desert-based judgments of blame before we invoke these judgments as a reason for punishment. In contrast, if our burden of proof jurisprudence authorizes us to acquit a defendant on the basis of a considerably lower confidence level than the one required to convict him, we should be allowed to invoke desert to limit punishment based on a lower confidence level than that which is required when we invoke desert to affirmatively justify punishment.

What remains, then, is to determine what degree of confidence we can justifiably have regarding whether humans have the amount of control that is necessary to ground non-instrumental judgments of desert and blame. Put differently, how sure can we reasonably be that we have the kind of free will retributivists assume we have? As the discussion in Part III demonstrates, we cannot be too sure that we have free will or that we have the kind of control that basic desert requires. If the brightest scientific and philosophical minds of our generation vehemently disagree about whether we have the kind of freedom that grounds judgments of blame and praise, it stands to reason that we are not justified in having a high degree of confidence in the existence of such freedom.

If we are not justified in believing that we have free will with a very high degree of confidence, we should not be invoking desert as an

144. See *In re Winship*, 397 U.S. 358, 361 (1970).

145. See *Martin v. Ohio*, 480 U.S. 228, 230 (1987).

affirmative reason for punishment, at least if we agree with Blackstone that it is best that ten guilty people go free than that one innocent be convicted.¹⁴⁶ Because we care more about wrongful convictions than about mistaken acquittals, we ought to proceed with humility when we look for affirmative reasons to impose punishment. If we are not convinced to a high degree of confidence that a certain belief tied to our punishment theories is justified, we should not be invoking that belief as a justification to punish. Given the current state of the free will debate, we should not have very high confidence in the belief that desert-based judgments are justified. For this reason, we should abstain from invoking retribution as a reason for punishing.

The same does not apply when desert serves as a limit on punishment. When invoking desert as a limiting principle, we are invoking it as a reason to abstain from causing suffering rather than as a reason for inflicting it. The degree of confidence that we should have in the belief that desert judgments are justified should be lower when we intend to use these judgments as a source of exculpation than when we intend to use them as a source of incrimination. I previously argued that parents should feel more comfortable rewarding their children on the basis of uncertain information than punishing them on the basis of such information.¹⁴⁷ If this is right, I would argue that, *mutatis mutandi*, our uncertainty regarding whether desert-based judgments are justified counsels in favor of allowing such judgments to play a role in exculpating defendants but abstaining from invoking such judgments to punish them.

V. CONCLUSION

Pursuant to the theory of “selective incompatibilism” that I advocate for in this Article, retribution ought to serve as a limit on punishment, but not as a justification for its imposition. While mixed theories of punishment are typically open to the “cherry-picking” objection that Zaibert leveled against Parfit’s brand of compatibilism, I believe that selective compatibilism is equipped to meet these objections. Mixed theories of punishment are often attacked as arbitrary or ad-hoc because most seem to be the product of desperate attempts to avoid the unappealing outcomes that consequentialist approaches to punishment seem to countenance, such as punishing the innocent or imposing the death penalty for speeding. But selective incompatibilism is not motivated by such concerns. Instead, selective incompatibilism is the

146. See BLACKSTONE, *supra* note 127.

147. See *supra* Part IV.B.

2019]

SELECTIVE INCOMPATIBILISM

1001

product of a dispassionate assessment of the free will debate and the implications that this debate ought to have for punishment theory.

While compatibilists and incompatibilists about free will surely feel confident about their views, an unbiased assessment of the debate reflects that the conversation about free will has reached a dialectical stalemate. The implications that follow from this are twofold. First, we should be wary of invoking free will or the related concepts of desert and blame as reasons for deliberately inflicting suffering on sentient creatures. If people who have been thinking about free will and desert for the entirety of their professional lives cannot agree on whether we have free will or whether desert judgments are justified, the criminal justice system should not place free will or desert at the center of punishment theory. Second, while the pervasive and deep disagreement about free will should lead us to abstain from invoking desert as a reason for punishment, it should not lead us to refuse to invoke desert as a limit on punishment. Because there is an asymmetry between inculcation and exculpation, we are justified in using desert as a reason against punishment but not as a reason in favor of its imposition. Since the mixed theory of punishment that follows from selective incompatibilism is grounded on the uncontroversial asymmetry between inculcation and exculpation, I would hope that whatever objections Zaibert musters against selective incompatibilism, cherry-picking is not one of them.