Michigan Law Review

Volume 90 | Issue 8

1992

Justifiably Punishing the Justified

Heidi M. Hurd

University of Pennsylvania

Follow this and additional works at: <https://repository.law.umich.edu/mlr>

Part of the [Constitutional Law Commons](#), [Public Law and Legal Theory Commons](#), and the [Rule of Law Commons](#)

Recommended Citation

Heidi M. Hurd, *Justifiably Punishing the Justified*, 90 MICH. L. REV. 2203 (1992).

Available at: <https://repository.law.umich.edu/mlr/vol90/iss8/2>

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

JUSTIFIABLY PUNISHING THE JUSTIFIED

Heidi M. Hurd*

INTRODUCTION	2204
I. THE DILEMMA	2205
A. <i>The Correspondence Thesis</i>	2205
B. <i>Role-Relative Morality: The Rejection of the Correspondence Thesis</i>	2212
II. THE PRESUPPOSITIONS OF THE DILEMMA	2217
A. <i>The Asymmetry of Law and Morality</i>	2217
B. <i>The Limited Authority of Law</i>	2222
III. SEIZING THE FIRST HORN: A DEFENSE OF PERSPECTIVALISM	2224
A. <i>Practical Errors: Pragmatic Foundations for Perspectivalism</i>	2226
1. <i>The Argument from Personal Error</i>	2226
2. <i>The Argument from Example</i>	2227
3. <i>The Argument from Opportunity</i>	2230
4. <i>Evaluating the Error Thesis</i>	2233
B. <i>The Rule-of-Law Values: Principled Foundations for Role-Relative Morality</i>	2236
1. <i>The Protection of Individual Liberty</i>	2238
2. <i>The Protection of Reliance Interests</i>	2240
3. <i>The Protection of Equality</i>	2252
C. <i>The Values of Democracy and the Separation of Powers: Principled Foundations for Constitutional Perspectivalism</i>	2255
1. <i>The Classic Arguments for Structural Pluralism</i> ..	2256
2. <i>The Rule-of-Law Values Revisited</i>	2274
3. <i>The Argument from Institutional Efficiency</i>	2275

* Assistant Professor of Law, University of Pennsylvania. B.A. 1982, Queen's University; M.A. 1984, Dalhousie University; J.D. 1988, Ph.D. Candidate, A.B.D. 1992, University of Southern California. — Ed. This paper has benefited substantially from the thoughtful comments that I received from those who attended its colloquia presentations at the Northwestern University School of Law, the Vanderbilt University School of Law, the University of Pennsylvania Law School, and the University of Iowa Philosophy Department. Very special thanks go to Frank Goodman for the rigorously detailed and challenging reply that he gave to the piece during the University of Pennsylvania Legal Studies Workshop. I am also grateful for the extensive comments provided by Larry Alexander, Daniel Epstein, William Ewald, Richard Fumerton, Barbara Herman, Scott MacDonald, Michael Moore, and Edwin Smith.

IV. SEIZING THE SECOND HORN: A DEFENSE OF THE	
CORRESPONDENCE THESIS	2279
A. <i>Consequentialism and the Correspondence</i>	
<i>Thesis</i>	2280
1. <i>Rightmaking Consequences for</i>	
<i>Consequentialism</i>	2280
2. <i>The Truth of the Correspondence Thesis</i>	2293
3. <i>Limiting Cases: Self-Referential Rightmaking</i>	
<i>Consequences</i>	2295
B. <i>Deontology and the Correspondence Thesis</i>	2301
1. <i>Sporting Competitions</i>	2303
2. <i>Gladiators and Innocent Shields</i>	2307
3. <i>Self-Defense</i>	2308
4. <i>Status-Based Relationships and Contractual</i>	
<i>Roles</i>	2312
CONCLUSION	2321

[A]ll law is universal but about some things it is not possible to make a universal statement which will be correct. In those cases, then, in which it is necessary to speak universally, but not possible to do so correctly, the law takes the usual case, though it is not ignorant of the possibility of error. . . . When the law speaks universally, then, and a case arises on it which is not covered by the universal statement, then it is right, when the legislator fails us and has erred by over-simplicity, to correct the omission — to say what the legislator himself would have said had he been present And this is the nature of the equitable, a correction of law where it is defective owing to its universality. . . . [W]hen the thing is indefinite the rule also is indefinite, like the lead rule used in making the Lesbian moulding; the rule adapts itself to the shape of the stone and is not rigid, and so too the decree is adapted to the facts.

—Aristotle, *Nicomachean Ethics*¹

INTRODUCTION

Contemporary moral philosophy, political theory, and jurisprudence have converged to create a quite baffling dilemma. This dilemma is generated by the apparent incompatibility of three principles, each of which grounds features of our system of law and government, and each of which carries substantial normative weight. The first I shall call the punishment principle — a moral principle, doctrinally entrenched in American criminal and civil law, which holds that individuals who are morally justified in their actions ought not to be

1. ARISTOTLE, *Nicomachean Ethics*, in 2 THE COMPLETE WORKS OF ARISTOTLE 1729, 1795-96 (David Ross trans., revised by J. Urmson, Jonathan Barnes ed., 1984).

blamed or punished for those actions. The second is the principle of the rule of law — a complex jurisprudential principle that requires law to conform to a set of formal values, such as generality and coherence, as a means of protecting substantive moral values like liberty and equality. The third is the principle of democracy and the separation of powers — a principle of political morality that vindicates the right of majorities to be self-governing by assigning policymaking powers to a democratic legislature and restricting the executive and judiciary to the secondary tasks of policy implementation and application.

These three principles now serve as cornerstones of our legal and political systems. Yet, if genuinely incompatible, one or more of them must be abandoned: we must resign ourselves to the punishment of the justified or sacrifice systemic values that have long been invoked to justify our commitment to structural pluralism and rule-governed adjudication.

I. THE DILEMMA

Because our fundamental moral, political, and jurisprudential principles are not in obvious conflict, the dilemma engendered by their mutual defense takes some construction. Let me begin at its seemingly remote beginning.

A. *The Correspondence Thesis*

Consider the following hypothetical. Smith is attacked by a hoodlum while walking her dog through the city park. Smith justifiably believes that her life is in peril, and she is thus forced to choose between killing the hoodlum or being killed or maimed herself. Jones is a jogger who witnesses the hoodlum's attack on Smith. Jones must choose between permitting Smith to kill the hoodlum or intervening to prevent that killing. Long is a concession stand owner who also witnesses the event. Long must choose between restraining Jones from intervening to prevent the hoodlum's death or allowing that intervention.

The morality of each actor's choice appears to be determined by what I will call the *correspondence thesis*. The correspondence thesis asserts a moral claim about the justification of codependent actions. It holds that the justifiability of an action determines the justifiability of permitting or preventing that action. According to the correspondence thesis, if Smith is justified in killing the hoodlum (as a means of self-defense), then Jones is not justified in intervening to prevent that killing, and hence Long is justified in restraining Jones' intervention.

The correspondence thesis rests on the intuition that because an action cannot be simultaneously right and wrong, one actor cannot be justified in performing an act while another is simultaneously justified in preventing that performance.² The intuitive plausibility of the thesis can be cashed out as follows. Step one: Right action is action that accords with the balance of reasons for action.³ Reasons for action are objective in the sense that their rightmaking characteristics are universal. If it is right for one to do an act, it must be right for all others that one do it. If, for Smith, the balance of objective reasons for action favors action *A* (where *A* is killing the hoodlum in self-defense), then the balance of reasons in favor of Smith's doing *A* must be the same for Jones and Long, i.e., it must dictate the conclusion on their part that it is right that Smith do *A*. Step two: Where other actors face choices between alternative actions that will or will not thwart action *A*, the rightness of *A* entails the wrongness of those actions that will thwart *A*. Jones, for example, faces the choice between intervening to prevent Smith from killing the hoodlum or not intervening. It would be morally anomalous if *A* were the right thing for Smith to do, while preventing *A* was the right thing for Jones to do. The objective reasons that constitute the content of morality must make it right (for Jones as for everyone else) that Smith do *A*, so those same reasons seemingly cannot make it right that Smith not do *A*. Hence, the balance of reasons for action must demand the permission of justified actions and the restraint of attempts to thwart justified actions.

It is useful at this early stage to stop to consider a series of objections that might be made to the correspondence thesis. By putting

2. As Kant argued:

Each member of the commonwealth has rights of coercion in relation to all the others, except in relation to the head of state. For . . . he alone is authorised to coerce others without being subject to any coercive law himself. . . . But if there were two persons exempt from coercion, neither would be subject to coercive laws, and neither could do to the other anything contrary to right, which is impossible.

IMMANUEL KANT, *On the Common Saying: "This May Be True in Theory, But It Does Not Apply in Practice,"* in KANT'S POLITICAL WRITINGS 61, 74-75 (H.B. Nisbet trans., Hans Reiss ed., 1970) (emphasis added).

3. Joseph Raz puts this principle as follows: "It is always the case that one ought, all things considered, to do whatever one ought to do on the balance of reasons." JOSEPH RAZ, *PRACTICAL REASON AND NORMS* 36 (1975) (footnote omitted). As Raz makes clear, the phrase *ought all things considered* functions in this principle to indicate "what ought to be done on the basis of all the reasons for action which are relevant to the question, and not only on the basis of the reasons the agent in fact considered or could have considered." *Id.* at 36 n.*. For similar statements of the conditions of right action, see STEPHAN L. DARWALL, *IMPARTIAL REASON* 99 (1983); JOHN RAWLS, *A THEORY OF JUSTICE* 341, 408 (1971). For alternative conceptions that epistemically limit the conditions of right action, see RICHARD A. FUMERTON, *REASON AND MORALITY: A DEFENSE OF THE EGOCENTRIC PERSPECTIVE* 90-128 (1990); Donald Davidson, *How is Weakness of the Will Possible?*, in *MORAL CONCEPTS* 93 (Joel Feinberg ed., 1970). I criticize such alternative conceptions *infra* text accompanying notes 151-55.

these to rest at this point, we can both better appreciate the claim made by the correspondence thesis and forestall future confusion about its implications. First, one might be tempted to resist the correspondence thesis because of the following sort of case. Suppose that Green has grounds for reasonably believing that she is being stalked by a man who seeks to kill her. Suppose further that she reasonably believes that a man who she discovers quietly circling her house is the stalker. She thus reasonably believes her life to be in peril and shoots at the man as he approaches her. However, the man, Brown, is there simply to read her electric meter. Reasonably believing his life to be in peril, Brown shoots back, misses, and is killed by another shot from Green.

Is this not a case of two individuals, each of whom is entitled to prevent the other from doing a justified act? Does this case not demonstrate that the correspondence thesis is false, because the justifiability of Green's conduct does not make it unjustifiable for Brown to resist it, and the justifiability of Brown's conduct does not make it unjustifiable for Green to resist it? Each appears to be justified in attempting to thwart the actions of the other. Therefore, the justifiability of one actor's conduct does not appear to determine the justifiability of the other's permission of that conduct.

This analysis is initially plausible only because the meaning of the phrase *justified action* is equivocal. The phrase is alternatively employed to capture both *right* action and *nonculpable* action. But the conditions of right action and the conditions of nonculpable action are distinct. And the correspondence thesis is a thesis solely about the conditions of right action; it is not a thesis about the conditions of nonculpable action. Let me explain.

An individual is nonculpable if she does an action justifiably believing it to be right. Alternatively, an individual is culpable if she does an action believing it to be wrong or unjustifiably believing it to be right. Culpability is thus a condition of an actor's state of mind. It reflects the degree of epistemic justification with which an actor concludes that her actions are right. An actor is epistemically justified in believing an action to be right if, under the circumstances, she has invested a reasonable amount of time, talent, diligence, and resources to acquire information about her circumstances and to determine the demands of morality in such circumstances. She has invested a reasonable amount of time, talent, diligence, and resources in these tasks if the costs of investing more of these goods are greater than the costs of a wrong decision discounted by its probability. In the case described above, one might well conclude that Green justifiably believed herself entitled

to shoot Brown. She rightly believed that an innocent person is entitled to use deadly force in cases of imminent peril, and, given what she knew of the activities of her stalker, she reasonably concluded that she was in such peril and, thus, that the costs of investing further time to determine more accurately the identity of the unknown man in her garden were sufficiently great as to outweigh the discounted costs of shooting an innocent man. At the same time, Brown justifiably believed that he was entitled to defend himself. He rightly concluded that, as an innocent person, he was entitled to use deadly force to defend himself against imminent peril, and he invested a reasonable amount of time and diligence to determine that Green was indeed placing him in such peril.

The correspondence thesis is thus plainly false if construed as a thesis about the conditions of culpability. Two individuals can justifiably believe that they are each entitled to thwart the actions of the other, for they can possess sufficiently different amounts of information, time, talent, and resources as to make it reasonable for them to entertain different beliefs about what they are each entitled to do. Construed epistemically, the thesis that the justifiability of an action determines the justifiability of permitting or preventing that action is thus false.

But the correspondence thesis is not an epistemological thesis about the conditions of culpability. It is a metaphysical thesis about the conditions of right action. While culpability is a function of an actor's state of mind, the rightness of an action is a function of the degree to which an action satisfies the objective criteria specified by our best normative theory.⁴ The correspondence thesis claims that the criteria of right action specified by our best moral theory cannot make contradictory actions simultaneously right. While two individuals might be justified in *believing* that thwarting one another's actions is right, it cannot in fact *be* right for each to thwart the actions of the other. If the action of one individual is right, it cannot be simultaneously right for the other to prevent that action.

Thus, while Green and Brown may have reasonably believed that they should shoot one another, only one of them acted rightly in so doing. Because Brown was not attacking Green, Green did not act rightly in shooting Brown. Had she refrained from shooting Brown, an innocent life would have been saved at no cost to her own life. If taking an innocent life is categorically wrong (according to our best

4. See *supra* note 3. For a critique of this distinction, see FUMERTON, *supra* note 3, at 102. I reply to criticisms of the sort raised by Fumerton *infra* text accompanying notes 151-55.

deontological theory), then Green acted wrongly in shooting Brown, for Brown was an innocent actor. If maximizing the preservation of innocent lives is right (according to our best consequentialist theory), then Green acted wrongly, for her action took one innocent life without saving others (for her life would have been preserved had she not shot Brown). Once it is clear that the correspondence thesis is not a thesis about culpability, but a thesis about right action, it is clear that hypotheticals such as that involving Green and Brown are not counterexamples to it.

However, one might think that, while the previous hypothetical of Green and Brown makes the correspondence thesis plausible for both deontological and consequentialist accounts of right action, the truth of the thesis is in fact theory-dependent. While the thesis might be necessarily true of a consequentialist account of right action, it need not be true of a deontological account of right action. I shall have a great deal to say about this suggestion in Part IV, which is devoted to determining the relative truth of the correspondence thesis for consequentialists and deontologists. At this stage, however, it is worth recognizing the plausibility of the thesis for both sorts of theorists.

Consequentialists are committed to the claim that right action consists in maximizing good consequences or minimizing bad consequences.⁵ Monistic consequentialists embrace a single-valued theory of the good. Thus, utilitarians define the good in terms of human pleasure or happiness.⁶ Egoists define the good in terms of what will serve the individual's own interests.⁷ And virtue theorists define the good in

5. For the classic defenses of consequentialism, see JEREMY BENTHAM, AN INTRODUCTION TO THE PRINCIPLES OF MORALS AND LEGISLATION 11-33 (Jonathan Barnes & H.L.A. Hart eds., 1970) (1781); JOHN S. MILL, *Utilitarianism*, in SELECTED WRITINGS OF JOHN STUART MILL 243 (Maurice Cowling ed., 1968); HENRY SIDGWICK, THE METHODS OF ETHICS 381-473 (1874). For more modern defenses, see FUMERTON, *supra* note 3, at 50-60, 90-223; DONALD REGAN, UTILITARIANISM AND CO-OPERATION (1980); PETER SINGER, PRACTICAL ETHICS 8-13 (1979); R.M. Hare, *Ethical Theory and Utilitarianism*, in 4 CONTEMPORARY BRITISH PHILOSOPHY 113 (H.D. Lewis ed., 1976); Kai Nielson, *Against Moral Conservatism*, 82 ETHICS 219 (1972); Peter Railton, *Alienation, Consequentialism, and the Demands of Morality*, 13 PHIL. & PUB. AFF. 134 (1984). For critical discussions of consequentialism, see DAVID LYONS, FORMS AND LIMITS OF UTILITARIANISM (1965); THE LIMITS OF UTILITARIANISM (Harlan B. Miller & William H. Williams eds., 1982); DEREK PARFIT, REASONS AND PERSONS 24-51 (1984); ANTHONY QUINTON, UTILITARIAN ETHICS (1973); SAMUEL SCHEFFLER, THE REJECTION OF CONSEQUENTIALISM (1982); UTILITARIANISM AND BEYOND (Amartya Sen & Bernard Williams eds., 1982); Bernard Williams, *A Critique of Utilitarianism*, in UTILITARIANISM: FOR AND AGAINST 75 (J.J.C. Smart & Bernard Williams eds., 1973).

6. See, e.g., BENTHAM, *supra* note 5, at 11, 34-50; MILL, *supra* note 5, at 249; SIDGWICK, *supra* note 5, at 30-42.

7. See, e.g., Jesse Kalin, *In Defense of Egoism*, in ETHICAL THEORY: CLASSICAL AND CONTEMPORARY READINGS 85 (Louis P. Pojman ed., 1989); FUMERTON, *supra* note 3, at 129-67. For helpful critical discussions of ethical egoism, see KURT BAIER, THE MORAL POINT OF VIEW 188-91 (1958); DAVID GAUTHIER, MORALS BY AGREEMENT (1986); MORALITY AND RATIONAL SELF-INTEREST (David P. Gauthier ed., 1970); THOMAS NAGEL, THE POSSIBILITY OF

terms of what will make individuals virtuous.⁸ Pluralist consequentialists, on the other hand, embrace a multivalued theory of the good. They maintain that the good should be maximized but then define the good in terms of some complex combination of the above sorts of values.

As a thesis about the conditions of right action, the correspondence thesis appears necessarily true for consequentialists.⁹ In our previous tale in which a hoodlum attacks Smith while Jones, the jogger, and Long, the concessionaire, look on, the death of the hoodlum is a better state of affairs than is its only alternative, the death of Smith. And it is a better state of affairs not only for Smith, but also for Jones and Long. Because the only criterion of right action for the consequentialist is whether an action promotes good states of affairs, it follows that Smith, Jones, and Long should all promote this good state of affairs. At the very least, it cannot be right on consequentialist grounds for Jones to interfere with Smith's act of self-defense, even if for some reason it is not obligatory for Jones to join Smith in her endeavor.

In contrast to consequentialists, deontologists are committed to the claim that the goodness of an act lies not in its consequences but in the inherent quality of the act itself.¹⁰ According to deontological moral

ALTRUISM (1970); PARFIT, *supra* note 5, at 3-24; JAMES RACHELS, *THE ELEMENTS OF MORAL PHILOSOPHY* 65-78 (1986); SIDGWICK, *supra* note 5, at 72-79, 107-74; W.D. FALK, *Morality, Self, and Others*, in *ETHICS* 349 (Judith Thomson & Gerald Dworkin eds., 1968); ALASDAIR MACINTYRE, *Egoism and Altruism*, in *2 THE ENCYCLOPEDIA OF PHILOSOPHY* 462 (Paul Edwards ed., 1967); Brian Medlin, *Ultimate Principles and Ethical Egoism*, 35 *AUSTL. J. PHIL.* 111 (1957).

8. See, e.g., ARISTOTLE, *supra* note 1; ALASDAIR MACINTYRE, *AFTER VIRTUE* 210-15 (1981); BERNARD MAYO, *ETHICS AND THE MORAL LIFE* (1958). For general discussion and critical examinations of virtue-based theories of the good, see LAWRENCE C. BECKER, *ON JUSTIFYING MORAL ARGUMENTS* 181-91 (1973); JOHN M. COOPER, *REASON AND HUMAN GOOD IN ARISTOTLE* (1975); WILLIAM K. FRANKENA, *ETHICS* 3-71 (2d ed. 1973); IRIS MURDOCH, *THE SOVEREIGNTY OF GOOD* 1-45 (1970); CHRISTINA SOMMERS, *VICES AND VIRTUES IN EVERYDAY LIFE* 143-231 (1985); RICHARD TAYLOR, *ETHICS, FAITH, AND REASON* (1985); *THE VIRTUES* (Robert B. Kruschwitz & Robert C. Roberts eds., 1987); JAMES D. WALLACE, *VIRTUES AND VICES* (1978); Robert B. Loudon, *On Some Vices of Virtue Ethics*, 21 *AM. PHIL. Q.* 227 (1984); Gregory E. Pence, *Recent Work on Virtues*, 21 *AM. PHIL. Q.* 281 (1984); Gregory W. Trianosky, *Supererogation, Wrongdoing, and Vice: On the Autonomy of the Ethics of Virtue*, 83 *J. PHIL.* 26 (1986); Susan Wolf, *Moral Saints*, 79 *J. PHIL.* 419 (1982).

9. See *infra* section III.A for a much more extensive analysis of this supposition.

10. For the classic articulation of deontological moral theory, see IMMANUEL KANT, *CRITIQUE OF PRACTICAL REASON* (Lewis W. Beck trans., Bobbs-Merrill Co. 1956) (1978); IMMANUEL KANT, *GROUNDWORK OF THE METAPHYSIC OF MORALS* (H.J. Paton trans., 1964); IMMANUEL KANT, *LECTURES ON ETHICS* (Louis Infield trans., Methuen & Co. 1979) (1930). For more contemporary discussions, see H.B. ACTON, *KANT'S MORAL PHILOSOPHY* (1970); BAIER, *supra* note 7, at 187-213; C.D. BROAD, *FIVE TYPES OF ETHICAL THEORY* 116-42 (1930); ALAN DONAGAN, *THE THEORY OF MORALITY* (1977); C.E. HARRIS, *APPLYING MORAL THEORIES* 127-57 (1986); THOMAS NAGEL, *MORTAL QUESTIONS* (1979); D.D. RAPHAEL, *MORAL PHILOSOPHY* 55-66 (1981); DAVID ROSS, *KANT'S ETHICAL THEORY* (1954); ROBERT P. WOLFF, *THE AUTONOMY OF REASON: A COMMENTARY ON KANT'S "GROUNDWORK OF THE META-*

theories, certain act-types are intrinsically right or intrinsically wrong. Thus, if it is right to preserve innocent life, one cannot take an innocent life even if, in so doing, one saves a great many more innocent lives.¹¹ Contemporary deontologists often refine this anticonsequential claim by describing morality as a set of agent-relative categorical imperatives.¹² Morality directs each agent not to kill innocent persons even if such a killing would prevent a greater number of similar killings by others.

A deontological, or *agent-relative*, theory of morality does not share the characteristic that appears to make the correspondence thesis necessarily true for the consequentialist. It is logically possible that a (certain sort of) deontological theory would render the correspondence thesis false.¹³ Suppose, for example, that morality contained an agent-relative obligation not to kill and an agent-relative permission (or obligation) to kill when necessary to preserve one's own life. Under such a set of moral directives, it would be right for Smith to defend herself by killing her attacker. But suppose that this morality also contained an agent-relative obligation to prevent killings by others, even when those killings are in self-defense. Then it would be right for Jones to prevent Smith from (rightly) defending herself. On such a theory, there is no correspondence between what it is right for Smith to do and what it is right for Jones to do.

While such a deontological theory entails the rejection of the correspondence thesis, it does not contain any formal contradiction. Yet its content is highly implausible. Agent-relative morality is still universal: if it is right for Smith to defend her own life, it is right for Jones to defend his own life. The view here considered has to both

PHYSICS OF MORALS" (1973); Philippa Foot, *Morality as a System of Hypothetical Imperatives*, 84 PHIL. REV. 305 (1972).

11. As Finnis puts it, "one should not choose to do any act which *of itself does nothing but* damage or impede a realization or participation of any one or more of the basic forms of human good." JOHN FINNIS, *NATURAL LAW AND NATURAL RIGHTS* 118 (1980).

[This] perhaps unfamiliar formulation . . . is well recognized, in other formulations: most loosely, as "the end does not justify the means"; more precisely, though still ambiguously, as "evil may not be done that good might follow therefrom"; and with a special Enlightenment flavour, as Kant's "categorical imperative": "Act so that you treat humanity, whether in your own person or in that of another, always as an end and never as a means only." *Id.* at 122 (quoting IMMANUEL KANT, *FOUNDATIONS OF THE METAPHYSICS OF MORALS* 47 (Lewis W. Beck trans., 1959) (1785)). Finnis has further developed his deontological theory in JOHN FINNIS, *FUNDAMENTALS OF ETHICS* (1983); JOHN FINNIS ET AL., *NUCLEAR DETERRENCE, MORALITY AND REALISM* (1987).

12. See, e.g., THOMAS NAGEL, *EQUALITY AND PARTIALITY* 40, 85-86 (1991); THOMAS NAGEL, *THE VIEW FROM NOWHERE* 152-54 (1986); NAGEL, *supra* note 7, at 90-95; PARFIT, *supra* note 5, at 143.

13. But see *infra* section III.B for a defense of the claim that a plausible deontological theory would incorporate the correspondence thesis.

admit this and contend that it is also right for Jones to prevent Smith from defending her life. The agent-relativity of such a deontological morality prevents these norms of rightness from contradicting themselves, but it does not prevent their mutual assertion from being highly untenable. Were morality to combine such maxims, individuals would be persistently prevented from doing what they would be entitled, and perhaps obligated, to do. Morality would make us moral gladiators in an arena in which one actor's moral success would depend upon another's moral failure.

It thus seems that any plausible deontological moral theory would subscribe to the correspondence thesis. If true, then the correspondence thesis does not depend for its intuitive viability on any particular moral theory.

B. *Role-Relative Morality: The Rejection of the
Correspondence Thesis*

Compelling as the correspondence thesis appears on any plausible moral theory, its intuitive appeal seems to give way when one is forced to take into account role-related considerations. Suppose now that Smith is married to a man who frequently beats her and her three children. She rightly believes that there is a strong probability that her husband will eventually attempt to kill them all. She also rightly believes that at that point she will be unable to defend any of them. And she rightly believes herself unable to guarantee their safety if they attempt to flee from him. She thus plausibly takes the balance of reasons for action to favor a preemptive strike, notwithstanding the fact that she recognizes and takes seriously the law's refusal to recognize a special battered spouse defense to homicide. She waits until her husband is asleep, enters his bedroom and fatally shoots him.

Suppose that Jones is a judge who must decide whether to convict or acquit Smith. Jones considers Smith morally justified in shooting her husband, for Jones concurs that the balance of reasons applicable to her choice favored such a killing. But Jones is a judge charged with the task of applying the law, and the law does not permit the acquittal of those who claim self-defense but admit the absence of any imminent threat of harm. Jones, like Smith, must thus decide whether to break the law. He must decide whether to comply with the decision rule that demands punishment of those who kill without legal justification or to break that rule and acquit one who is *morally* justified.¹⁴

14. I have found it useful throughout this article to invoke the distinction between conduct rules and decision rules earlier coined by Bentham and recently revived by Meir Dan-Cohen. See JEREMY BENTHAM, A FRAGMENT ON GOVERNMENT AND AN INTRODUCTION TO THE PRINCIPLES OF MORALS AND NATURE (1789) (Oxford: Clarendon Press, 1988).

Suppose, finally, that Long is responsible for designing and maintaining our political and legal systems. Among the many tasks required to foster democracy, preserve the separation of powers, maintain law and order, and ensure right action by citizens and officials is the task of appointing and disciplining members of the judiciary. Long, like Jones, recognizes the plausibility of Smith's claim of justification. Long also recognizes that Jones might rightly conclude that he should acquit Smith, notwithstanding the decision rule that requires her punishment. Long thus faces the choice between disciplining Jones, should Jones fail to comply with the decision rule, and licensing the sort of judicial legislation that justified judicial disobedience would represent.

We are now in a position to appreciate the dilemma to which I referred in the Introduction. Jones and Long face choices which pit our fundamental moral, political, and jurisprudential values against one another. If the correspondence thesis applies to acts of punishment in the same way that it applies to preventative and permissive acts, then Jones should acquit Smith, and Long should not discipline Jones for such an acquittal. It would seem that the correspondence thesis should be as true of acts of punishment as of acts of prevention, because punishment labels an act as wrong and thus serves to prevent future actors (including the actor who is punished) from performing that act in similar circumstances. If Smith's act is right, it seems wrong to respond to that act with sanctions that imply either that Smith should not have performed that act or that future actors should not perform the same act in similar circumstances. But if judges like Jones acquit those who are morally justified in breaking the law, what will become of the rule of law? For law seemingly ceases to be law if judges are entitled to rethink its wisdom in every case to which it applies and to disregard it whenever it is inferior to the rule that they would fashion.¹⁵ And if system designers like Long refuse to discipline members of the judiciary who disobey democratically enacted decision rules, what will become of democracy and the separation of powers? For the majority does not rule itself, and the powers of gov-

PLES OF MORALS AND LEGISLATION 430 (Wilfrid Harrison ed., 1948); Meir Dan-Cohen, *Decision Rules and Conduct Rules: On Acoustic Separation in Criminal Law*, 97 HARV. L. REV. 625, 625-30 (1984); see also DAVID DAUBE, *FORMS OF ROMAN LEGISLATION* 24 (1956) (discussing distinction of conduct and decision rules in Talmudic law). Conduct rules constitute those rules that are intended to guide the daily behavior of citizens. Decision rules constitute those rules that are intended to guide judges in the adjudication of disputes concerning the behavior of citizens.

15. As Lon Fuller argued, the substitution of individual judgment constitutes a failure to comply with the "internal morality of the law," and thus "results in something that is not properly called a legal system at all. . . ." LON FULLER, *THE MORALITY OF LAW* 39 (2d ed. 1969).

ernment are not kept separate, if individuals are entitled to substitute their own considered opinions for those of a democratic legislature.¹⁶

Yet the correspondence thesis applies only if the balance of reasons for action is indeed the same for Jones and Long as for Smith. That is, it applies only if it is right for Jones and Long, as well as for Smith, that Smith violate the conduct rule prohibiting homicide in a situation in which there is no fear of imminent harm. The rule of law is secure if reasons apply to Jones that do not apply to Smith — reasons that make Smith's conduct wrong for Jones, notwithstanding its rightness for her. Such reasons, if sufficiently weighty, might justify the judicial punishment of an admittedly justified offender. Even if such judicial reasons are insufficient to justify Jones' punishment of Smith, our commitment to democracy and the separation of powers is secure if reasons apply to Long that do not apply to Jones — reasons that make Jones' acquittal of Smith wrong for Long, notwithstanding its possible rightness for Jones. In such a case, Long might be justified in punishing Jones for not punishing Smith.

But the cost of preserving our systemic values is the cost of abandoning the punishment principle. According to this principle, individuals whose actions are morally justified ought not be punished for their actions.¹⁷ This principle could be justified by what legal philosophers and criminal law theorists have called a "weak version" of the principle of retribution. The principle of retribution holds that individual desert is both a necessary and a sufficient condition of praise and blame: morally culpable individuals deserve punishment and thus should be punished even when their punishment will not serve to advance any welfare goal, for example, the prevention of crime (the strong version of the retributive principle¹⁸); morally innocent individuals (individuals who are justified, excused, or who otherwise fail to satisfy conditions of responsibility) do not deserve punishment and thus should not be punished, even when their punishment might fur-

16. [G]ood reasons for avoiding the creation of nondemocratic political elites militate against judges being given the authority to modify extant legal rules on the basis of their personal perceptions of what is likely to have the best consequences in the instant case. . . . [T]he whole point of the legislative enterprise would be lost if the courts were given the authority to subvert it.

ROLF E. SARTORIUS, *INDIVIDUAL CONDUCT AND SOCIAL NORMS* 178 (1975).

17. "[E]very moral man must be appalled at the judicial execution of the innocent or at the punishment, torture, and killing of the innocent. Indeed, being appalled by such behavior partially defines what it is to be a moral agent." Nielson, *supra* note 5, at 220.

18. See, e.g., John Rawls, *Two Concepts of Rules*, 64 *PHIL. REV.* 3, 4 (1955) ("What we may call the retributive view is that punishment is justified on the grounds that wrongdoing merits punishment."). For a powerful defense of this position, see Michael S. Moore, *The Moral Worth of Retribution*, in *RESPONSIBILITY, CHARACTER AND THE EMOTIONS: NEW ESSAYS IN MORAL PHILOSOPHY* 179 (Ferdinand Schoeman ed., 1987).

ther some welfare goal (the weak version of the retributive principle¹⁹).

More interesting for present purposes is the support that the correspondence thesis provides for the punishment principle. If the correspondence thesis is true of acts of punishment, then the justifiability or unjustifiability of an individual's act determines the justifiability or unjustifiability of that individual's punishment, for the balance of reasons that determines the morality of her act is identical to the balance of reasons that determines the morality of her punishment. Insofar as an actor is justified in performing an act, another actor cannot be justified in punishing her for that act.²⁰

The punishment principle is intuitively compelling because it appears morally repugnant to blame or punish individuals who have done precisely what they should have done. To abandon the punishment principle is to threaten individuals with a catch-22: to escape punishment they must depart from the balance of reasons for action — and so act immorally; to preserve the morality of their conduct they must conform to the balance of reasons for action — and so suffer punishment.

But the recognition of role-relative reasons for action puts the punishment principle in jeopardy, for it entails the indefensibility of the correspondence thesis. If morality requires conduct in accord with the balance of reasons for action, and if that balance differs among the citizen, judge, and institution designer, then there may be instances in which the punishment of the justified is itself justified. The most that the punishment principle can do under role-relative morality is serve

19. See, e.g., Rawls, *supra* note 18, at 11.

20. The vindication of the punishment principle by the correspondence thesis distinguishes the punishment principle from what Rolf Sartorius has called the *reflection principle*. According to the reflection principle, "[w]here an individual has correctly decided that he ought to do X, any higher-order judgment about his decision to do X or his actual act of doing it ought to license or approve of, rather than disapprove of or penalize, the decision and/or the act itself." SARTORIUS, *supra* note 16, at 56-57; see also Larry Alexander, *Law and Exclusionary Reasons*, 18 PHIL. TOPICS 5, 10-11 (1990) [hereinafter Alexander, *Law*]; Larry Alexander, *Pursuing the Good — Indirectly*, 95 ETHICS 315, 323-25 (1985) (discussing Sartorius' reflection principle) [hereinafter Alexander, *Pursuing the Good*]. The reflection principle, as stated by Sartorius, presupposes that the reasons for action may differ as between an actor and one who judges that actor. It thus constitutes a normative thesis about the sort of judgment that ought to be made concerning an actor's conduct, rather than a conceptual thesis about the sort of judgment that must be made about that conduct. If the correspondence thesis is true, the reflection principle is also true (and it thus collapses into the punishment principle). But this is because the reasons for higher-order judgments about an actor's conduct are identical to the reasons for that conduct; hence, if the conduct is justified, the only justifiable judgment about that conduct must be one of approval. It is not because the reasons for judgment include a second-order principle that holds that, notwithstanding the fact that the reasons for conduct and the reasons for judgment about that conduct may not be different, the reasons for judgment ought to be calculated *as if* they were identical to the reasons for conduct.

as a reason, albeit a possibly weighty one, to refrain from punishing the justified offender.²¹ But because such a reason must be added to the balance of reasons for and against punishment, it may be insufficient to compel the acquittal of the justified.²²

The dilemma is now clear. Either we must give up the punishment principle, and acquiesce in the punishment of justified offenders, or we must give up our traditional understanding of the systemic principles that currently justify our system of adjudication. More precisely, (1) either judges must violate the punishment principle and punish citizens who justifiably break the law, or they must violate the rule of law and disobey decision rules that require the punishment of the justified; similarly, (2) either system designers must violate the punishment principle and punish judges who justifiably disobey decision rules (by refusing to punish justifiably disobedient citizens), or they must violate the principle of democracy and the separation of powers by permitting the substitution of individual judicial judgment for the will of a legislative majority.

The first alternative in each of these cases appears the inevitable result of embracing the possibility of role-relative morality, or what I shall call *perspectivalism*.²³ If institutional roles create new reasons (or

21. The punishment principle, under this analysis, collapses into the reflection principle. See *supra* note 20.

22. I say this on the assumption that the punishment principle (like the reflection principle with which it is identical on this analysis, see *supra* notes 20-21) cannot function as an exclusionary reason for action. This assumption follows from the arguments that I have advanced on behalf of the claim that exclusionary reasons are incoherent. For a summary of these arguments, see *infra* text accompanying notes 39-40. For the original, quite lengthy version of these arguments, see Heidi M. Hurd, *Challenging Authority*, 100 YALE L.J. 1611, 1620-41 (1991).

If the deontologist considers both the punishment principle and the role-relative reasons for action unique to citizens, judges, and institution designers to be both intrinsically rightmaking and at least practically incompatible, then the deontologist will be confronted with moral conflict that only the sort of moral balancing described here can seemingly resolve. For a classic discussion of deontological balancing, see W.D. ROSS, *THE RIGHT AND THE GOOD* 16-47 (1930). But cf. Barbara Herman, *Obligation and Performance: A Kantian Account of Moral Conflict*, in *IDENTITY, CHARACTER AND MORALITY: ESSAYS IN MORAL PSYCHOLOGY* 311, 319 (Owen Flanagan & Amélie O. Rorty eds., 1990); Barbara Herman, *The Practice of Moral Judgment*, 82 J. PHIL. 414 (1985) [hereinafter Herman, *Moral Judgment*] (arguing that practical reason requires not balancing of reasons, but detecting interlocking exceptions to reasons). For an appreciation of the general debate concerning the existence and resolution of deontological moral conflicts, see the essays collected in *MORAL DILEMMAS* (Christopher W. Gowans ed., 1987).

23. It is crucial not to confuse a theory of role-relative morality with a theory of metaethical relativism. A theory of role-relative morality is entirely consistent with a theory of metaethical realism — that is, a theory that holds that propositions about morality are true or false independent of anyone's beliefs about them. Indeed, the puzzle with which this article deals is only interesting if it is supposed that metaethical relativism is false. If metaethical relativism were true, the correspondence thesis would be trivially false. The truth of moral propositions would be relative to the beliefs of individuals (in the case of metaethical subjectivism) or to the beliefs of communities (in the case of metaethical conventionalism). Insofar as beliefs can differ, the morality of any given act could differ as between individuals or communities. Hence, it would be trivially true that an actor might be morally justified (from her perspective) in doing an act that another (who

eliminate otherwise valid reasons) for action, the correspondence thesis is false and morality may compel the punishment of the justified. The second alternative appears the inevitable result of denying the possibility of role-relative morality. If citizens share with officials an identical set of reasons for action, then the correspondence thesis is true, and justifiable disobedience by a citizen will justify disobedience by an official, which in turn will justify approval by a system designer.

II. THE PRESUPPOSITIONS OF THE DILEMMA

The dilemma that I have sketched would not arise if either of two things were true. First, if the content of the law perfectly mirrored the content of morality, so that citizens and officials were never called upon to do anything contrary to the balance of moral reasons for action, then citizens and officials would never be justified in violating the law. Second, if the law, *qua* law, provided reasons for unconditional obedience, then, even if it did not perfectly mirror morality, it would nevertheless preclude any instances of justified disobedience by citizens. In either case, we would have no cause for moral concern over the punishment of disobedient citizens, and hence in either case there would be no occasion on which a judge might be justified in refusing to punish a disobedient citizen. The dilemma therefore is a product of two presuppositions: (1) that the content of law departs from the content of morality; and (2) that the authority of law departs from the authority of morality.

A. *The Asymmetry of Law and Morality*

The content of law fails to cohere with the content of morality even in legal systems, like our own, that make real efforts to pass laws that accord with the demands of morality. Because this is not obviously true, and because the dilemma with which I am concerned presupposes its truth, it is useful to see why it is true. Let us begin with the criminal law proper and then turn to civil sanctions under tort law.

Anglo-American penal systems purport not to punish justified of-

occupies a different role) is morally justified (from his perspective) in punishing. Therefore, the puzzle generated in this Part is only significant if metaethical relativism is false. I have elsewhere argued that it is. Heidi M. Hurd, Note, *Relativistic Jurisprudence: Skepticism Founded on Confusion*, 61 S. CAL. L. REV. 1417, 1459-506 (1988).

While this article thus presupposes that metaethical relativism is false, it does not presuppose that metaethical realism is true. Some noncognitivists believe that they can duplicate virtually all of the claims of metaethical realists. See SIMON BLACKBURN, *SPREADING THE WORD* (1984); Jeremy Waldron, *The Irrelevance of Moral Objectivity*, in *NATURAL LAW THEORIES* 158 (Robert P. George ed., 1992). If they are right, then the question of whether morality is role-relative is as significant for them as for moral realists.

fenders.²⁴ Sometimes this is captured by the fact that many of those who are justified in their actions are not interpreted to be "offenders" at all. The sheriff who arrests a federal mail carrier for murder, while that carrier is carrying the mail, *literally* "obstructs the mails," but does not *legally* obstruct the mails, because his obstruction is justified. The sheriff who risks damage to federal property in his attempt to arrest a murderous mail carrier does not *recklessly* risk its damage because legal recklessness amounts to taking a substantial *and unjustifiable* risk.²⁵ In these cases, the good consequence achieved (apprehending a murderer) renders the sheriff's conduct nonpunishable, for the sheriff is said to lack either the *actus reus* or the *mens rea* of the offense in question. Sometimes the refusal to punish such a justified offender is captured by a separate defense, such as the Model Penal Code's "balance of evils" defense: Conduct that is otherwise criminal is not punishable if "the harm or evil sought to be avoided by such conduct is greater than that sought to be prevented by the law defining the offense charged"²⁶

Yet, despite the manifest reluctance of the criminal law to punish justified offenders, it will nevertheless demand such punishment in the cases with which I will be concerned in this article. In such cases, the law demands action that fails to comport with the balance of reasons. It nevertheless preempts individual calculation of or action on that balance. Consider four such cases. First, individuals who violate laws that they rightly consider fundamentally mistaken will find themselves without recourse to exculpatory doctrines, their moral calculations preempted by those of the lawmaker. Thus, in the event that abortion again becomes illegal, women who rightly conclude that the law prohibiting abortion fails adequately to weigh their liberty interests against the interests of the fetuses they carry will nevertheless be thought to be preempted by the law from deciding that they should have abortions.

Second, civilly disobedient offenders, who rightly believe that a law is so immoral that it justifies illegal actions to obtain its repeal, are thought to be rightfully punished. Draft protesters, for example, might rightly think that a war is sufficiently immoral that it justifies them in trespassing into Selective Service offices to disrupt operations.

24. JOSHUA DRESSLER, UNDERSTANDING CRIMINAL LAW 249-51 (1987); JEROME HALL, GENERAL PRINCIPLES OF CRIMINAL LAW 444-45 (1947).

25. These illustrations are loosely constructed from the facts of *United States v. Kirby*, 74 U.S. (7 Wall.) 482 (1868) (county sheriff prosecuted for obstructing the mail after arresting on-duty federal mail carrier suspected of murder). See *infra* text accompanying note 88.

26. MODEL PENAL CODE § 3.02 (Proposed Official Draft 1962).

But the moral justification invoked by these protesters for violating laws prohibiting criminal trespass and the destruction of government property will not be afforded legal recognition; rather, it will be held to be preempted by the laws that license the draft.²⁷

Third, our law requires judges to punish offenders for whom the law has created an exception but who fail the conditions of that exception. The battered wife might rightly think that if she does not kill her husband in his sleep she will be unable to prevent him from taking her life. But she fails to fulfill the conditions of self-defense, for she does not bring about her husband's death while under a threat of imminent harm. In such a case, she may not be civilly disobedient (in the classic sense), for she may believe that the law is (in general) justified in requiring imminent peril as a condition for the use of deadly force, and she probably does not act in order to bring about any changes in that law. Nevertheless, as in cases of civil disobedience, the law of self-defense preempts her ability to invoke a *legal* justification for her (morally) justified conduct.²⁸

The fourth class of justified offenders who are unable to avail themselves of any legal justifications consists of those offenders whose justification for violating the law is too personal to be given legal recognition by the balance of evils defense, or by other exculpatory doctrines. The citizen who finds herself at a red light in the middle of the night while in the middle of nowhere may rightly conclude that, while the inconvenience of stopping is negligible, the danger to herself and others from running the red light is nil. Such a citizen may well be justified in running the red light, but the law will regard her justification as preempted by the rules of the road.

In each of these cases, Anglo-American criminal law demands the punishment of justified offenders. Now consider the civil law. Unlike

27. As section 3.02 of the Model Penal Code states:

(1) Conduct which the actor believes to be necessary to avoid a harm or evil to himself or to another is justifiable, provided that:

...

(c) a legislative purpose to exclude the justification claimed does not otherwise plainly appear.

MODEL PENAL CODE § 3.02 (Proposed Official Draft 1962); see *State v. Tate*, 505 A.2d 941 (N.J. 1986) (holding necessity defense unavailable to quadriplegic who possessed marijuana for medical purposes because state statute allowing use of controlled substances under physician supervision suggested a legislative intent to exclude the defense).

28. Thus, according to section 3.02 of the Model Penal Code:

(1) Conduct which the actor believes to be necessary to avoid a harm or evil to himself or to another is justifiable, provided that:

...

(b) neither the Code nor other law defining the offense provides exceptions or defenses dealing with the specific situation involved . . .

MODEL PENAL CODE § 3.02 (Proposed Official Draft 1962).

the criminal law, the civil law generally does not "punish" justified offenders with civil liability. This is made particularly clear by negligence law, which explicitly exempts from liability those who act on the balance of reasons for action. One must pay damages under the law of negligence only if one is unjustified in having taken risks that materialized in harm to a plaintiff.

Tort law is less clear about its refusal to "punish" the justified in cases that invite the application of strict liability. Insofar as tort law imposes liability in these cases without regard to the justification with which a defendant harmed the plaintiff, it might be thought to "punish" the justified. Thus, for example, in cases like *Vincent v. Lake Erie Transportation Co.*,²⁹ individuals are thought to be justified in using or damaging the property of others, for doing so prevents a greater loss to themselves. Yet the law requires that they compensate those whom they justifiably injure. In cases that involve harms caused by privately owned wild animals, owners are thought to be justified in keeping the animals so long as they do so nonnegligently. Yet the law holds them liable for all injuries that their animals cause.³⁰ In cases that involve injuries caused by abnormally dangerous activities, such as dynamiting tunnels or transporting gasoline, those who engage in such activities are thought to be justified if they conduct themselves nonnegligently. But they are nevertheless held liable for all injuries that they cause.³¹

On their face, these sorts of cases certainly seem to suggest that in at least some instances individuals will be civilly "punished" for morally justified conduct. But this suggestion is misleading. While the balance of reasons for action may favor the use of others' property in circumstances of necessity, or the ownership of wild animals, or abnormally dangerous activities, the balance of reasons for action in such cases probably also requires the compensation of those injured. Thus, while defendants in such cases may have justifiably caused harm, they are probably unjustified in their refusal to abide by the civil law that requires compensation for such harm. Put differently, the reasons that judges would have to exact compensation in these cases are reasons for defendants to volunteer compensation in the first place. For in such cases, compensation vitiates the defendant's unjust enrichment. If redressing unjust enrichment is a reason for a judge to require compensation, it is a reason for a citizen to offer it.³² And defendants who in

29. 124 N.W. 221 (Minn. 1910) (holding shipowner liable for loss of plaintiff's dock even though shipowner acted reasonably in tying ship to dock during storm).

30. RESTATEMENT (SECOND) OF TORTS § 509 (1965).

31. RESTATEMENT (SECOND) OF TORTS §§ 519, 520 (1965).

32. The thesis that unjust enrichment constitutes a morally legitimate reason for the imposi-

fact *wrongly* refuse to compensate those whom they injure are not *justified* offenders of the civil law. Hence, while judges should require them to compensate those whom they injure, this imposition of liability does not mean that genuinely justified offenders will be held civilly liable. For genuinely justified offenders are those who would be justified both in causing harm and in not paying for that harm.

Tort law thus generally does not make genuinely justified offenders liable. The only exceptions to this occur when strict liability is imposed on defendants who neither acted negligently nor were personally enriched by conduct that caused harm to a plaintiff. Some product liability cases have this flavor,³³ as do some cases that involve abnormally dangerous activities pursued for public (rather than private) benefit.³⁴ In such cases, strict liability is imposed on individuals who have done the right thing and who have not been unjustly benefited by so doing. Unless morality provides individuals with a reason to pay for all the harms that they cause, these parties seem to have no moral reason to offer compensation to those whom their justified conduct injures. Insofar as tort law imposes liability anyway, it "punishes" the justified.

It is thus contingently true that our legal system sometimes imposes criminal and civil liability on actors who act morally.³⁵ We

tion of civil liability has been advanced by JULES COLEMAN, *MARKETS, MORALS AND THE LAW* 166-201 (1988); JEFFRIE G. MURPHY & JULES L. COLEMAN, *THE PHILOSOPHY OF LAW: AN INTRODUCTION TO JURISPRUDENCE* 167-89 (1984); Jules L. Coleman, *Moral Theories of Torts: Their Scope and Limits: Part II*, 2 *LAW & PHIL.* 5 (1983); George P. Fletcher, *Fairness and Utility in Tort Theory*, 85 *HARV. L. REV.* 537 (1972). See also Heidi M. Hurd, *Correcting Injustice to Corrective Justice*, 67 *NOTRE DAME L. REV.* 51, 62-84 (1991) (discussing Jules Coleman's unjust enrichment theory of tort liability).

33. *Koos v. Roth*, 652 P.2d 1255 (Or. 1982); *Seigler v. Kuhlman*, 502 P.2d 1181 (Wash. 1973), *cert. denied*, 411 U.S. 983 (1974).

34. *Brody v. Overlook Hosp.*, 296 A.2d 688 (N.J. Super. Ct. Law Div. 1972), *revd.*, 317 A.2d 392 (N.J. Super. Ct. App. Div. 1974), *aff'd.*, 332 A.2d 596 (N.J. 1975).

35. Some would go further. Some would claim that it is necessarily true that in all legal systems the content of law must depart from the content of morality. Hence, it is necessarily true that the law will call for the punishment of the justified.

Fred Schauer, for example, has recently argued that legal rules are *necessarily* suboptimal in comparison to moral rules. FREDERICK SCHAUER, *PLAYING BY THE RULES* 100-02 (1991). Legal rules are, for Schauer, a species of prescriptive rules authored to guide the behavior of others. *Id.* at 17-37. Prescriptive rules are probabilistic generalizations that are inevitably over- and underinclusive. *Id.* at 31-34. Not all dogs create annoying disturbances, and many things that are not dogs do create such disturbances; hence, a rule prohibiting dogs from public buildings, while probabilistically related to the justification for that rule, is simultaneously over- and underinclusive. As Schauer maintains, "it is logically impossible for a rule to generate a result for a particular case superior to the result that would have been generated in the absence of rules, but . . . it is indeed quite possible for rules to generate results in particular cases that are inferior to those generated without them." *Id.* at 101. He explains:

In many cases, indeed in most cases, the result indicated by applying a rule will be the same as the result indicated by directly applying the rule's background justifications But . . . there will be some cases in which the result indicated by the rule will be inferior to the result

come, then, to the second presupposition of the dilemma that I sketched in the opening Part — the presupposition that the law lacks the authority to provide citizens with an overriding moral reason for obedience.

B. *The Limited Authority of Law*

The most influential contemporary view of the law's authority, that advanced and defended by Joseph Raz, conceives of law as a practical authority capable of issuing exclusionary reasons for obedience.³⁶ According to this view, the sheer fact that a law has been enacted provides individuals with a reason for obedience that trumps (virtually) all reasons for disobedience.³⁷ In the face of such an exclusionary reason for obedience, the right thing to do, all things considered, is to

indicated by direct application of its justification. Yet these instances will not be accompanied by any offsetting instances in the opposite direction.

... The only standard for measurement would be the one ultimate justification, and thus the result indicated by that justification would by definition be the best for any particular case. Consequently, any rule-indicated result diverging from the justification-indicated result, would *eo ipso* be an inferior result.

Id. at 100-01.

Schauer's claim that legal rules must necessarily depart from moral rules depends on the assumption that the function of legal rules is to guide action. From this assumption, Schauer reaches his necessity claim in two steps: (1) For legal rules to guide action more effectively than do moral rules, legal rules must be clearer, simpler, and more accessible than are the rules of morality; (2) to possess these features, legal rules must be over- and underinclusive vis-à-vis the more complex, shaded, reasons for action that morality provides.

I shall not pursue whether Schauer is correct either in assuming that the function of law is to guide conduct or in calculating that law must necessarily depart from morality in order to fulfill that function. It is sufficient for our purposes to note that actual legal systems do in fact punish some justified offenders because such systems lack a legal defense for every case in which a disobedient citizen has a moral justification. It follows from this fact alone that, unless the law itself provides citizens with an overriding moral reason to obey the law, there will be cases in which individuals are morally justified in violating the law.

36. JOSEPH RAZ, *THE AUTHORITY OF LAW* 3-33 (1979) [hereinafter RAZ, *AUTHORITY*]; JOSEPH RAZ, *THE MORALITY OF FREEDOM* 23-109 (1986) [hereinafter RAZ, *MORALITY*]; RAZ, *supra* note 3.

37. This description of Raz's conception of the dynamics of practical authority is deliberately simplistic. Raz argues that the law of a state obligates only if and when the state constitutes a *legitimate* practical authority acting within its proper jurisdiction. The imposition of these constraints makes Raz's theory both sophisticated and discriminating, for it makes the question of whether the law is authoritative a piecemeal one, the answer to which varies from individual to individual and circumstance to circumstance. While I have argued that Raz's theory succumbs to the same problems that plague simpler accounts of practical authority, see Hurd, *supra* note 22, it does not do so because it simplistically asserts that the law unexceptionally obligates. Rather, it does so because it asserts that the law has the power sometimes to obligate. For an excellent collection of articles analyzing and critiquing Raz's theory of practical authority, see Symposium, *The Works of Joseph Raz*, 62 S. CAL. L. REV. 731 (1989); see also RICHARD E. FLATHMAN, *THE PRACTICE OF POLITICAL AUTHORITY* 109-25 (1980) (critiquing Raz's concept of the authority of rules); LESLIE GREEN, *THE AUTHORITY OF THE STATE* 36-51 (1988) (following Raz's theory of authority); NEIL MACCORMICK, *LEGAL RIGHT AND SOCIAL DEMOCRACY* 232-46 (1982) (employing Raz's understanding of the authority of law to analyze whether law is necessarily coercive); Alexander, *Law*, *supra* note 20 (critiquing Raz's theory of legal authority).

comply with the law.³⁸

Notwithstanding the explanatory power of this model, it has come under considerable attack by those who distrust the claim that the law should be thought to preempt individual moral judgment. I have argued, for example, that exclusionary reasons for action are conceptually incoherent and, hence, that a theory that accords law practical authority cannot be defended.³⁹ Others have maintained that a system of law premised on a claim of practical authority is inherently immoral.⁴⁰ Still others have maintained that as a psychological matter, law does not affect our practical reasoning in a manner that is compatible with an exclusionary account of its authority.⁴¹

Those who have shared my rejection of a theory of practical authority have defended a version of one of two alternative models that accord the law only limited authority. One model postulates that law possesses what I have called "influential authority": it provides citizens and officials with a new reason for obedience, but it does not ex-

38. More accurately, Raz argues that a legitimate practical authority provides "protected reasons" for action. RAZ, *AUTHORITY*, *supra* note 36, at 18. A protected reason is complex. First, it embodies a first-order "content-independent reason" to act as commanded by the law. A first-order content independent reason for action stems from the sheer fact that an authority has issued a request or command. As Raz describes it:

A reason is content-independent if there is no direct connection between the reason and the action for which it is a reason. The reason is in the apparently "extraneous" fact that someone in authority has said so, and within certain limits his saying so would be reason for any number of actions, including (in typical cases) for contradictory ones.

RAZ, *MORALITY*, *supra* note 36, at 35; see also H.L.A. HART, *ESSAYS ON BENTHAM* 254-55 (1982) (coining the notion of content-independent reasons employed by Raz); Joseph Raz, *Voluntary Obligations and Normative Powers*, 46 *ARISTOTELIAN SOC. PROC.* 79, 95-98 (Supp. 1972) (explicating concept of content-independent reasons).

In addition to providing a new first-order content-independent reason to act as commanded, a protected reason for action also gives a second-order exclusionary reason of the sort described in the above text. Hence, upon being commanded to act by a practical authority, one has both a new reason to act as commanded *and* a reason to refrain from acting upon the reasons that one previously had not to act as commanded. A protected reason for obedience to the law thus makes obedience rational, all things considered, for the only reasons to consider are those that require obedience (including, and perhaps limited to, the first-order content-independent reason provided by the protected reason itself). For a more extensive discussion of the components of a protected reason for action, see Hurd, *supra* note 22, at 1616-20.

While the terminology of protected and exclusionary reasons is distinctive to Raz and the literature about Raz, the notion that there are rights, rules, and performative acts (e.g., promises) that prohibit us from giving weight to certain types of reasons for action (e.g., efficiency, social utility) is commonly invoked in anti-utilitarian philosophy. See, e.g., ROBERT NOZICK, *ANARCHY, STATE, AND UTOPIA* 28-35 (1974) (discussing rights as "side-constraints"); RAWLS, *supra* note 3, at 150-61 (arguing that benefits of slavery are not to be considered in setting up a just society).

39. See Hurd, *supra* note 22, at 1614-39.

40. See Michael S. Moore, *Authority, Law, and Razian Reasons*, 62 *S. CAL. L. REV.* 827, 859-73 (1989).

41. See FLATHMAN, *supra* note 37, at 110-12; SCHAUER, *supra* note 35, at 88-93.

clude reasons for disobedience.⁴² According to this model, individuals are justified in breaking the law in instances in which they accurately calculate that the reasons for disobedience outweigh the reasons for obedience (including among those latter reasons the fact that the law demands obedience). The alternative model that I have defended takes the law to possess, at most, "theoretical authority": it cannot provide a new reason for action, but it may provide a new reason to believe that there are other reasons for action — reasons that existed antecedent to the law's enactment.⁴³ That is, the law may provide evidence of the conduct that morality requires. According to this model, individuals are justified in breaking the law whenever they rightly consider it to be poor evidence of their moral obligations.

Under either of these alternative models of the law's limited authority, the demands of practical rationality obligate individuals to weigh the merits of obedience against those of disobedience. In those instances in which individuals rightly determine that the law fails to reflect the balance of reasons for action, individuals are justified in breaking the law (and are indeed obligated to do so). This conclusion constitutes the second presupposition of the dilemma, for it provides a theoretical defense of justified disobedience and generates the need for a compatible theory of adjudication. Such a theory will be called upon either to embrace perspectivalism and punish morally justified offenders or to preserve the correspondence thesis and acquit morally justified offenders. The first alternative will require the rejection of the punishment principle. The second will require us to sacrifice our traditional understanding of systemic values such as the rule of law, democracy, and the separation of powers.

III. SEIZING THE FIRST HORN: A DEFENSE OF PERSPECTIVALISM

Those who have confronted the dilemma that I have sketched have felt compelled to seize its first horn and abandon the punishment principle. As Larry Alexander has maintained, "[t]here is an always-possible gap between what we have reason to do, all things considered . . . , and what we have reason to have our . . . officials . . . require us to

42. See FLATHMAN, *supra* note 37, at 119; SCHAUER, *supra* note 35, at 112-18; Moore, *supra* note 40, at 862-73.

43. See Hurd, *supra* note 22, at 1667-77; Heidi M. Hurd, *Sovereignty in Silence*, 99 YALE L.J. 945, 1007-28 (1990); see also Alexander, *Law*, *supra* note 20, at 7 ("Law in fact is not itself a reason to act, first-order or otherwise. Although law can affect our first-order reasons, it cannot be a first-order reason."); Donald H. Regan, *Authority and Value: Reflections on Raz's Morality of Freedom*, 62 S. CAL. L. REV. 995, 1003-18 (1989) (arguing that laws constitute "indicator rules" that sum up, but do not provide, first-order reasons for action).

do.”⁴⁴ Recognition of this gap, according to Alexander, requires that we give up the punishment principle and punish those who do the right thing in breaking the law. “[I]n one role we occupy, that of authority, we should impose sanctions on ourselves for actions that are correct in another role we occupy, that of subjects of rules. It may be morally good that we punish ourselves for breaking morally good rules for morally good reasons.”⁴⁵ Fred Schauer has similarly argued that authority is asymmetrical. “If the lack of a (good moral) reason for obeying authority does not entail the lack of a (good moral) reason for imposing it, then, *a fortiori*, the practice of imposing authority, and of enforcing compliance with rules *qua* rules, can be seen as sometimes justified”⁴⁶

Implicit in the willingness to reject the punishment principle is a commitment to role-relative morality or perspectivalism. The “always-possible gap” that Alexander identifies is a product of the view that public roles provide officials with reasons for action that are different from the reasons for action possessed by those who do not occupy such roles. Hence, a single act may be moral from the perspective of a citizen but immoral from the perspective of a judge. Alternatively, it may be moral from the perspective of a judge but immoral from the perspective of a system designer. I will ultimately argue that role-relative morality is misconceived and, hence, that we need not and should not embrace the paradoxical conclusion that morality justifies the punishment of the morally justified.⁴⁷ But it is important to appreciate the power of the claim that the dilemma is an inescapable one, and that our theoretical commitments suffer the least damage by forfeiting the punishment principle.

There are two potential sources of role-relative morality. First, role-relative morality is thought to follow from the need to prevent the moral errors that are the inevitable product of human fallibility. Theorists who consider error to be the sole source of role-relative morality provide what I take to be a pragmatic defense of the need to punish the justified. Such a defense furnishes a weak theory of perspectivalism, for, in the absence of moral error, there is no basis for defending role-relative morality, and hence no basis for abandoning the correspondence thesis. But role-relative morality is also thought by some to be a product of institutional concerns that would survive the (admittedly unlikely) perfection of our practical reasoning. Theorists who take

44. Larry Alexander, *The Gap*, 14 HARV. J.L. & PUB. POLY. 695, 695 (1991).

45. *Id.* at 696.

46. SCHAUER, *supra* note 35, at 129-30.

47. See *infra* Part IV.

these institutional concerns to be the source of role-relative morality provide a principled defense of perspectivalism that serves as a substantially stronger reason for sacrificing the punishment principle.⁴⁸ If such theorists are successful in advancing a principled defense of role-relative morality, then we must conclude that morality is inherently paradoxical. It requires us to do what it requires others to prevent or punish.

In this Part, I propose to explore these sources of role-relative morality. In the subsequent Part, I shall ask whether the considerations to which these sources give rise are genuinely role-relative.

A. *Practical Errors: Pragmatic Foundations for Perspectivalism*

Those who seek to derive role-relative morality from considerations of human fallibility must be prepared to explain how such considerations generate differential reasons for action for citizens, officials, and institution designers. At least three accounts appear to be available.

1. *The Argument from Personal Error*

Alexander has maintained that one reason for the gap between justified disobedience and justified punishment is that "we as the subjects of rules are fallible, and we are more likely to produce those consequences demanded by our moral principles if we are governed not directly by those principles but by blunt (over- and under-inclusive) rules that are relatively easy to follow and to monitor."⁴⁹ This insistence on rule-governed conduct recalls John Stuart Mill's conviction that "[w]hatever we adopt as the fundamental principle of morality, we require subordinate principles to apply it by . . ."⁵⁰ Insofar as the law provides just such a set of subordinate principles for action, this thesis suggests that compliance with the law will accomplish results that are morally superior to those achieved by attempts to comply with the dictates of morality.

This claim, by itself, fails to provide a defense of role-relative morality. It merely suggests that, by virtue of our inevitable fallibility, all of us, citizens, judges, and institution designers alike, have a reason to

48. For a discussion of the distinction between pragmatic and principled defenses of role-relative morality, see *infra* text accompanying notes 72-73.

49. Alexander, *supra* note 44, at 696; see SCHAUER, *supra* note 35, at 149-55.

50. MILL, *supra* note 5, at 266. Such a conviction was born of the recognition that "there is not time, previous to action, for calculating and weighing the effects of any line of conduct on the general happiness." *Id.* at 264. Absent such time, individuals are bound to make mistakes that lead to wrong action. For a similar defense of the argument from personal error, see Rawls, *supra* note 18, at 23.

defer to the law if the law prescribes conduct that accords with the relevant moral principles more often than does our own calculation of those principles. Those who contemplate disobedience (of either decision rules or conduct rules) must factor into their calculations the possibility that they are in error in thinking that the relevant rules are over- or underinclusive. If this calculation is itself likely to be erroneous, then concern for personal moral error may provide a good reason for general deference to the law. But such a reason appears equally applicable to citizens, judges, and institution designers.

To generate differential reasons for action, those who appeal to the potential for personal error must provide some account of why the fear of personal moral error should lead judges to comply with decision rules in instances in which the fear of personal moral error should not lead citizens to comply with conduct rules. Such an account would be available if judges were prone to greater error than citizens. Under such circumstances, a judge would have a more weighty reason to abide by the decision rule requiring punishment of a battered wife than the battered wife would have in deciding whether to violate the conduct rule prohibiting homicide in the absence of imminent peril. But because there are grounds for thinking that judges and other officials are at least as well situated as are citizens to assess the balance of reasons for action,⁵¹ the simple argument from personal error fails to provide a compelling reason to embrace role-relative morality.

2. *The Argument from Example*

A more sophisticated version of the error thesis is more persuasive. Alexander has sketched its foundations as follows:

Rules are formulated by finite, fallible human beings whose ability to foresee and consider possible applications of rules is limited. The rules they promulgate will therefore be imperfect. Suppose an agent correctly sees that the rule that commands him to do [not-]A in situation X would be improved were it to command him to do A in X instead. Suppose he then disobeys the rule and does A.

This argument assumes that A will necessarily have better consequences than [not-]A because a better rule would command A and not [not-]A. But . . . [s]ome agents will mistakenly believe that departures from rules are warranted when in fact no exception is called for, and the negative effects of these mistaken exceptions may outweigh the positive effects of justified exceptions.⁵²

This more sophisticated argument appeals not to the possibility that

51. I have discussed the conditions that must be met to establish this in Hurd, *supra* note 43, at 1010-15.

52. Alexander, *Pursuing the Good*, *supra* note 20, at 322.

an actor's decision to break the law will itself be erroneous, but to the possibility that an actor's otherwise correct decision will cause erroneous decisions by others.⁵³ Its claim is that, because morality requires action based on the balance of reasons, all things considered, and one of the reasons against breaking the law that one must seemingly consider is the possibility that one's (otherwise justified) disobedience will serve as a poor example to others, then morality compels the assessment of the precedential value of one's justified acts. Where that precedential value is great, an otherwise justified act of disobedience may be unjustified.

This more sophisticated claim, like its simpler counterpart, fails, by itself, to provide citizens and officials with differential reasons for action. It merely suggests that all actors who debate the merits of abiding by the law must discount the reasons for disobedience by the probability that such disobedience will encourage a disproportionate amount of unjustified disobedience by others. In order for this claim to provide a basis for role-relative morality, it must be supplemented with an account of why a judge might have a reason to punish a citizen who has *correctly* calculated that her (otherwise justified) disobedience will not encourage a disproportionately greater amount of unjustified disobedience.

The necessary corollary to the sophisticated error thesis is the quite plausible claim that the precedential effect of a citizen's (otherwise justified) disobedience is a less weighty reason for the citizen to comply with the law than is the precedential effect of a judge's (otherwise justified) disobedience. Because public officers occupy highly visible positions, and because citizens and other officials often take their conduct within these positions to define the extent of obedience required by the law, their disobedience provides a more powerful incentive for (unjustified) disobedience by others than does that of most private citizens.⁵⁴ Insofar as judicial disobedience is more likely than

53. See FRED FELDMAN, *INTRODUCTORY ETHICS* 97 (1978), for the thesis that setting an example constitutes a consideration that should enter into the balance of reasons for action.

54. [E]veryone, judges and citizens alike, has an obligation to support just institutions. This obligation to support just laws does not give rise to a general obligation to obey the laws on the part of *citizens*, . . . because "it is a melodramatic exaggeration to suppose that every breach of law" by a citizen sets such a bad example that it will endanger the just system of laws everyone has an obligation to support. . . . [I]t is not nearly such an exaggeration to claim that *judges* cannot disregard the laws as they exercise their offices without demotivating by their example a large number of their fellow citizens to comply themselves. For citizens might well think that if even judges can be lawless in their official behavior, what reason is there for ordinary citizens to be law-abiding? If this is so, then judges have a special obligation to obey the laws as they judge

Moore, *supra* note 40, at 836 (quoting RAZ, *MORALITY*, *supra* note 36, at 102) (footnotes omitted). This argument should be supplemented with the observation that, insofar as a court's precedent formally influences future judges, a court's disobedience appears formally to encourage

private disobedience to encourage moral mistakes by other citizens and officials, the argument from example purportedly provides a more weighty reason for judges to abide by decision rules than for citizens to abide by conduct rules. It provides a basis for thinking that a judge might be morally compelled to punish a battered wife who was morally justified in breaking the law of homicide.

There is a tempting rejoinder to the argument from example. This rejoinder draws on the intuition that we should each be entitled to do the right thing, even if it causes others to do the wrong thing. One way to unpack this intuition is to invoke an argument that limits the consequences that serve as reasons for and against an action to those that are the proximate result of that action. When another mistakenly employs our justified disobedience as precedent for an unjustified act of disobedience, that individual's violation is the sort of voluntary intervening act that breaks the causal chain between our conduct and any subsequent unjustified consequences. While all of the proximate consequences of our actions serve as reasons for and against those actions, consequences that are not proximate do not enter into the balance of reasons that determines right action. In short, while our error is of our concern, others' error is not, for the error of others is a proximate consequence of their voluntary acts, not ours.

As I will later argue, the attempt to invoke proximate cause limitations to curtail the consequences for which one is causally responsible is metaphysically untenable.⁵⁵ But, even if this rejoinder could be rendered metaphysically plausible, it would only provide us with a reason to think that those who set examples do not *cause* the unjustified conduct of those who take such examples to heart. This would be morally significant only if morality required us to refrain from causing harm but did not require us to act affirmatively to prevent harm when we are capable of doing so. While considerations of individual liberty may justify us in thinking that the law ought not to require good samaritanism,⁵⁶ those considerations are unlikely to be of a type that

future disobedience on the part of other courts. See RAZ, *AUTHORITY*, *supra* note 36, at 105-11. Insofar as courts will erroneously misconstrue examples of judicial disobedience, precedential disobedience may encourage more unjustifiably disobedient adjudication than justifiably disobedient adjudication.

55. See *infra* text accompanying notes 143-50.

56. See FRIEDRICH A. HAYEK, *THE CONSTITUTION OF LIBERTY* 133-47 (1960); FRIEDRICH A. HAYEK, *THE ROAD TO SERFDOM* (1944); JOHN HOSPERS, *LIBERTARIANISM: A POLITICAL PHILOSOPHY FOR TOMORROW* (1971); TIBOR R. MACHAN, *INDIVIDUALS AND THEIR RIGHTS* (1989); JOHN S. MILL, *On Liberty*, in *SELECTED WRITINGS OF JOHN STUART MILL*, *supra* note 5, at 121, 189-207; NOZICK, *supra* note 38, at ix; MURRAY N. ROTHBARD, *THE ETHICS OF LIBERTY* 161-88 (1982); MURRAY N. ROTHBARD, *FOR A NEW LIBERTY* 148 (1978); Richard A. Epstein, *A Theory of Strict Liability*, 2 J. LEGAL STUD. 151, 189-204 (1973); John Hospers, *What Libertarianism Is*, in *THE LIBERTARIAN ALTERNATIVE* 3, 13 (Tibor R. Machan ed., 1974). For

will convince us that morality does not require it. Liberty interests may be sufficiently hefty to outweigh reasons for affirmatively acting to rescue others from peril, but it would take a particularly virile strain of libertarianism to exclude reasons for aid altogether from the balance of reasons that justifies action.⁵⁷

There thus seems to be no means of denying that practical rationality requires individuals to consider the likelihood that they will induce, by example, unjustified conduct on the part of others. Insofar as judges are more likely than citizens to encourage unjustified disobedience by their justified disobedience, it appears that judges have a more weighty reason to abide by legal rules than do citizens. As such, the argument from example appears to provide a promising basis for vindicating the sort of role-relative morality that would justify the punishment of justified offenders. The judge ought to punish the battered wife for justifiably killing her husband, because if she does not, other citizens will resort to deadly force in unjustified circumstances, and other judges will unjustifiably acquit those citizens.

3. *The Argument from Opportunity*

The third version of the error thesis is a sophisticated hybrid of the first two versions. It holds that one who can affirmatively act to prevent or reduce morally bad consequences has a reason to do so.⁵⁸ To generate a defense of role-relative morality, this premise must be combined with some account of how officials have opportunities that are unavailable to citizens to prevent moral harm through their adherence to the law.

This account is plausibly grounded in the following three-step ar-

two quite good collections of essays on the libertarian thesis that liberty is unjustifiably compromised by good samaritan legislation, see *THE LIBERTARIAN ALTERNATIVE*, *supra*; *THE LIBERTARIAN READER* (Tibor R. Machan ed., 1982).

57. Many libertarians are libertarians in part because they consider charity, generosity, and kindness to be virtuous only if uncoerced. To preserve opportunities for moral action they insist that the state must not legislate virtue. They thus recognize moral duties of good samaritanism, but insist that such duties cannot be meaningfully fulfilled if coerced. See, e.g., MACHAN, *supra* note 56, at 162-63; NOZICK, *supra* note 38, at 167-73; Epstein, *supra* note 56, at 200, 203-04.

58. This proposition is less controversial than it may first appear. Moral and political theorists of virtually all stripes are likely to sign onto it once they recognize its limitations. The principle as stated is a moral one, not a political one, and hence it need not concern libertarians and liberals who object to the legislation of good samaritan obligations. Moreover, the principle holds that the opportunity for good samaritanism is but *a* reason for action, and hence the principle is perfectly compatible with the supposition that in many circumstances other reasons for action will override it, for example, reasons given by the fact that aid to others will substantially jeopardize the life, liberty, or property of the would-be good samaritan. In the face of these limitations, the principle simply affirms the moral value of charity — a value that few are likely to deny and, hence, a value that in certain circumstances (i.e., those that pose the opportunity for aid to others) functions as one, among many, reasons for action.

gument. First, the demands of practical reason compel citizens to act on the balance of reasons for action available to them.⁵⁹ Whenever citizens conclude that the law conflicts with the balance of reasons for action, including in that balance the reasons for obedience provided by the arguments from personal error and example, citizens should break the law.⁶⁰

But citizens are fallible. They often lack the ability to assess accurately the balance of reasons for action.⁶¹ What is worse, they lack the ability to assess when they lack the ability to assess the balance of reasons for action.⁶² As a result, citizens would abide by the correct balance of reasons for action more often if they could escape the demands of practical reason and simply follow the law on every occasion, including those occasions on which the law both is and appears to them to be patently wrong or even gravely immoral.⁶³

Thus, citizens face a predicament. Practical rationality dictates

59. Insofar as right action constitutes action based on the balance of *all* reasons for action, see *supra* note 3, and insofar as actors may lack the time, talent, resources, or foresight, to determine *all* reasons for action applicable to a given decision, the most that practical rationality can require is that actors act on the balance of reasons for action reasonably available to them. See *infra* text accompanying notes 151-55.

60. This is a conclusion to which all theorists who accord law less than practical authority are committed. See *supra* notes 39-41.

61. As Holly Smith has argued, actors often suffer from both motivational and cognitive handicaps. "There are . . . at least four cognitive handicaps that could prevent human agents from utilizing a variety of moral principles in actual decision-making: incapacity to comprehend the principle, lack of sufficient information to apply it, erroneous empirical beliefs, and limited ability to make the requisite calculations." Holly M. Smith, *Two-Tier Moral Codes*, in *FOUNDATIONS OF MORAL AND POLITICAL PHILOSOPHY* 112, 115 (Ellen F. Paul et al. eds., 1990).

62. "If we do not trust a decision-maker to determine *x*, then we can hardly trust that decision-maker to determine that this is a case in which the reasons for disabling that decision-maker from determining *x* either do not apply or are outweighed." SCHAUER, *supra* note 35, at 98.

63. "Formal rules, because they refer to a few, easily identified factual circumstances and require no great ability in abstract reasoning or empirical knowledge for their application, enable ordinary, fallible human beings to come closer to realizing their abstract political/moral principles than would direct application of those principles . . ." Larry A. Alexander, *Painting Without Numbers: Noninterpretive Judicial Review*, 8 DAYTON L. REV. 447, 460 (1983). But, as Raz has argued, in order maximally to reduce moral error, such formal rules must be given the sort of preemptive weight that is denied to them under theories of influential and theoretical authority. Unless individuals surrender their own judgment, they will not prevent as many errors as rule-following will prevent in a regime in which the rules reflect the balance of reasons for action more often than do individual judgments.

Suppose I can identify a range of cases in which I am wrong more than the putative authority. Suppose I decide because of this to tilt the balance in all those cases in favour of its solution. . . . This procedure will reverse my independent judgment in a certain proportion of the cases. . . . If, we are assuming, there is no other relevant information available then we can expect that in the cases in which I endorse the authority's judgment my rate of mistakes declines and equals that of the authority. In the cases in which even now I contradict the authority's judgment the rate of my mistakes remains unchanged, i.e. greater than that of the authority. This shows that only by allowing the authority's judgment to pre-empt mine altogether will I succeed in improving my performance and bringing it to the level of the authority.

RAZ, *MORALITY*, *supra* note 36, at 68.

that they act on the balance of reasons for action available to them and thus that they break the law when they judge the balance of reasons so to dictate. But citizens are sufficiently prone to error in assessing that balance (as well as in assessing the degree to which they should discount their judgments about that balance by the likelihood of their own error) that they would do better to follow the law "blindly."

Finally, those who occupy institutional roles have an opportunity to rescue citizens from this predicament by acting affirmatively to prevent or reduce moral error. This power stems in part from their unique ability to influence by example. But judges also have the power to levy sanctions on citizens who break the law, and a threat of sanctions gives citizens a new and often weighty subjective reason for obedience.⁶⁴ In most instances, such a reason will be sufficient to tip the balance of reasons for action in favor of compliance. Punishment thus provides a judicial vehicle for reducing moral error. Insofar as one has a reason to act whenever one has an ability to prevent immoral consequences, and insofar as the power of punishment provides judges with a unique ability to prevent such consequences, judges have a reason to abide by the law (and thus to punish all lawbreakers) that is not a reason for citizens to abide by the law.

This argument differs from the previous two versions of the error thesis because it focuses, not on the worry that disobedience will either constitute moral error (on one's own part) or cause moral error (on the part of others), but on the worry that disobedience will constitute a failure to prevent moral error. If citizens would do better to follow the law even in circumstances in which practical rationality would dictate that they disobey it, and if judges have at their disposal a means of providing citizens with a reason for obedience that is so weighty that in most circumstances it will be sufficient to prompt citizens to obey the law, then judges should employ that power to rescue citizens from moral peril. Similarly, if judges would do better to follow entrenched decision rules "blindly," but if the demands of practical reason constrain them in so doing, then institution designers should use their powers of discipline to supplement the balance of judicial reasons for action with the subjective reason for obedience that derives from the threat of impeachment. Hence, judges and institution designers have reasons for punishment that those who are punished do not have when deciding to disobey the law.

64. Such a reason functions as a content-independent reason for action. See *supra* note 38. It is thus akin to, but not identical with, the sort of content-independent reason for action that the law, *qua* law, is thought to give by theorists who attribute to law either practical or influential authority. See *supra* notes 36-38 and accompanying text.

4. *Evaluating the Error Thesis*

The error thesis, in either or both of its two more plausible forms, constitutes a basis for role-relative morality only if citizens do not have, as a reason for obedience, the fact that judges are likely to induce error or fail to prevent error if they are forced to adjudicate cases involving civil disobedience. This is an inquiry to which I shall return in Part IV.

At this point it is useful to appreciate both the implications and limitations of the error thesis. First, the error thesis explicitly encourages the use of legal institutions as tools of deceit. It claims that citizens will achieve morally superior results if they come to believe the false proposition that they should never break the law.⁶⁵ But the inculcation of this belief comes at the price of sacrificing the morally right result in instances in which that result cannot be achieved without undermining false faith in the law's ability to achieve moral results or without failing to exploit an opportunity to strengthen that faith.

There are a number of reasons, extensively explored and articulated by others, to doubt that morality is ever best achieved through deceit. One might argue, for example, that deceiving others is intrinsically immoral and thus that it cannot be justified.⁶⁶ Yet, as Holly Smith has argued, one typically thinks deceit immoral because it causes false beliefs, and false beliefs cause wrong acts.⁶⁷ But the deceit accomplished by punishing the justified purportedly causes individuals to hold more true beliefs (about what they ought to do) than they

65. As Alexander maintains, the error thesis provides a compelling moral reason to create institutions "that demand that we act *as if* their decisions were morally preemptive of all other first-order moral reasons. But . . . their decisions cannot in fact be morally preemptive." Alexander, *Law*, *supra* note 20, at 10. Moreover, it may provide a compelling moral reason "not only to establish institutions that make such demands, but also to teach that their demands are morally preemptive (though they are not)." *Id.* at 11.

Dan-Cohen has argued that the natural acoustic separation that exists between citizens and officials makes it both possible and desirable to enact conduct rules that depart from decision rules. Dan-Cohen, *supra* note 14. Citizens can be given conduct rules that do not contain any exceptions that might be misconstrued or exploited. Officials can be provided with decision rules that call upon them to acquit citizens in exceptional circumstances. Such a combination both maximizes rule-following and allows for fairness. To the extent that conduct rules diverge from decision rules, however, they function as lies to the public. Contrary to what the public is told, ignorance of the law may excuse, use of deadly force will be allowed in special circumstances, and duress can exonerate. The public is deceived because the conduct rules encourage it to believe otherwise.

66. See Smith, *supra* note 61, at 124. As Smith points out, this sort of objection is available only to a deontologist. Insofar as the deontological prohibition against deceit is inconsistent with the deontological requirement that actors act in accordance with other agent-relative prohibitions and prescriptions, "[w]e cannot conclude that the prohibition against deceit, even for a deontologist, shows that [punishing the justified] should not be accepted as the best solution to the Problem of Error." *Id.*

67. *Id.* at 125-26.

would otherwise hold, and hence to act morally more often than they would otherwise act. Thus, in order to base an argument against deceit on some claim that deceit produces harmful beliefs, one would have to argue that true beliefs about the principles that justify actions are more important than are true beliefs about how one should act.

Alternatively, one could claim that deceit is inherently immoral because it deprives an agent of her autonomy.⁶⁸ But the background thesis would have to be that autonomy is a function of choices based on true beliefs about justificatory principles of action rather than on true beliefs about how one should act. One who thinks that autonomy requires moral action will be unpersuaded by the claim that a system that produces more moral action than its alternatives jeopardizes autonomy.

Bernard Williams has argued that, even if deceit can be morally innocuous, its use by the state will inevitably result in governmental manipulation of the populace that is antidemocratic, and this result is not morally innocuous.⁶⁹ Members of the state (for example, judges) will be called upon by citizens (for example, through the legislative enactment of particular decision rules) to act on rules that the former know are in some instances either over- or underinclusive. In those instances, the true moral rules will compel them to be unresponsive to majority will and hence undemocratic in their methods of governance. John Rawls has alternatively suggested that any attempt to inculcate false moral beliefs (however benign the background intention) will violate what he calls *the publicity condition* — a formal constraint that invalidates moral principles that are undermined by their public articulation.⁷⁰ Rawls takes such a condition as axiomatic to a public conception of justice and so concludes that, because it could not be rejected within the original position, it cannot be rejected outside it.

All of these arguments, to the extent that they point to genuine problems with the use of deceit, suffer from the same problem. Unless they bear some exclusionary status,⁷¹ these arguments can at most serve as reasons for judges and institution designers to refuse to sacrifice moral results as a means of preserving a moral myth. They must therefore be weighed against the reasons to think the myth a valuable one to preserve. Thus, such reasons, while important to explore and weigh, provide no basis for thinking that the error thesis could not serve as a source of role-relative reasons for action.

68. *Id.* at 126-27; see Herman, *Moral Judgment*, *supra* note 22, at 431.

69. See Williams, *supra* note 5, at 138-39; Smith, *supra* note 61, at 121-22.

70. See RAWLS, *supra* note 3, at 133; Smith, *supra* note 61, at 122-23.

71. See *supra* note 38 and accompanying text.

There is, however, a more important reason to think that the error thesis provides a questionable foundation for a defense of role-relative morality.⁷² Implicit in the error thesis is the presupposition that law allows individuals to act morally more often than does morality itself, for individuals can know and correctly apply the law in instances in which they cannot be sure of and cannot easily apply the general principles of morality. Yet for the law to approximate morality more often than does individual practical reason, the law must be laid down by individuals (1) who themselves understand and can apply the general principles of morality, (2) who understand how individuals can be confused about the content or application of those general principles of morality in particular circumstances, and (3) who recognize how to overcome this confusion by translating the general principles of morality into particular prescriptions or prohibitions that will be understood and applied in a manner that (more often than not) produces actions identical to those that the general principles of morality would require. Individuals capable of substituting a list of particular prescriptions and prohibitions for the general principles of morality would also be capable of identifying instances in which those prescriptions and prohibitions are over- or underinclusive (for one could not understand how a rule could approximate the outcome prescribed by its background principles without understanding the circumstances in which the rule does not apply). But individuals possessed of this sort of knowledge would seemingly be capable of remedying the moral error of others by educating them to be equally sophisticated moral reasoners, rather than by manipulating them through deception. To put it bluntly, if lawmakers know enough to manipulate citizens to act morally, they must know enough to educate citizens so that they need not manipulate them.

There may remain reasons to think that deception is preferable to education. Considerations of economy might suggest that only an elite few should master the general principles of morality. The rest should act on the rules that are produced by that moral elite, even when doing so produces moral errors, for the costs of those errors are not as great as the costs that would accompany the universal mastery of the principles that would reveal the exceptions to those rules.

But to recognize that legal deception can practically be replaced by moral education is to recognize that role-relative morality, if premised

72. This is a variation of the argument advanced by Holly Smith against two-tiered moral codes that replace the complex principles of morality with simple moral rules — rules that collectively comprise what Henry Sidgwick called an “esoteric morality.” See SIDGWICK, *supra* note 5, at 452-53; Smith, *supra* note 61, at 128-32.

on the error thesis, can be of pragmatic interest only.⁷³ In the absence of error the correspondence thesis would apply, because the reasons for and against a citizen's actions would exhaust the reasons for and against judicial action. While error theorists may be content to make their modest pragmatic claim because of their (quite plausible) confidence in human fallibility, it strikes me that one could make a considerably more interesting and theoretically powerful claim on behalf of role-relative morality. Even if we were all infallible moral reasoners, such that we never misconstrued others' examples of disobedience or otherwise miscalculated the balance of reasons for action, there might still be principled grounds for thinking that institutional roles create, and are in part defined by, reasons for action that are not applicable to individuals who do not occupy such roles. If this is the case, then, even if citizens never made mistakes about when morality required their disobedience of the law, judges might still have moral reasons to punish them for their justified disobedience. And even if judges never made mistakes about when morality required the acquittal of justifiably disobedient citizens, institution designers might still have moral reasons to punish them for their justified failure to punish justified offenders. That is, if principled perspectivalism is defensible, then morality is inherently paradoxical, for morality in principle requires the punishment of those who act morally. The abandonment of the punishment principle will not constitute a "second-best" solution to the problems of human imperfection. It will constitute part of the content of an ideal morality.

B. *The Rule-of-Law Values: Principled Foundations for Role-Relative Morality*

Lon Fuller maintained that a system can accomplish the rule of law only if its enactments are general, public, prospective, clear, logically consistent, practically possible, relatively constant, and predictably applied.⁷⁴ With the exception of the first (which simply states a

73. That is, the theory of role-relative morality is a strategic one, not a normative one. It holds that for practical reasons, not for theoretical reasons, judges ought to punish the justified. Just as one must recognize that one cannot aim directly at a target when shooting an arrow in a high wind, so judges must recognize that they cannot acquit the justified when adjudicating cases in a system in which those cases are erroneously construed as examples to the unjustified. But just as a high wind does not *alter* the location of the target one is aiming at but merely changes the way in which one aims at it, so too error does not *change* the content of the principles of morality but simply alters the practical manner with which those principles are best realized. If one's concern is with the content of morality, and not with the strategic methods necessary for its realization, one will find the thesis of role-relative morality to be beside the point, if that thesis is premised solely on the fact that actors make errors.

74. See FULLER, *supra* note 15, at 33-94.

formal condition for the very existence of a system of rules), these requirements speak to the importance of three values: the promotion of individual liberty; the protection of reliance interests; and the preservation of equality.⁷⁵

Many have claimed that these rule-of-law values "generally define the job of judging."⁷⁶ They "justif[y] the judiciary having a limited role in a democracy such as ours. . . . [They] mandate that judges should not dispense justice in some ad hoc, case-by-case basis."⁷⁷ Implicit in this claim is the view that rule-of-law values are reasons for action unique to adjudicators. They enter into the balance of reasons that justifies judicial action but not into the balance of reasons that justifies private conduct. As such, the rule-of-law values serve as a source of role-relative morality that might well justify the punishment of justified offenders.

In this section, I shall articulate the nature of these values and the reasons to think them role-relative. While these values are well known in the jurisprudential literature, their ability to provide principled foundations for role-relative morality is variable, and hence their separate treatment is necessary.

Throughout this discussion it is important to assume that error is not a concern. This assumption will allow us to establish whether there are any role-relative reasons for action that are principled rather than pragmatic. Our task is to determine whether morality is in principle paradoxical — whether it issues different reasons for action to different actors so that the same action might be justified, given the reasons for action applicable to one actor, and justifiably punished, given the reasons for action applicable to another. If the rule-of-law values depend for their normative force on the need for judges to prevent erroneous decisionmaking on the part of citizens (or other judges), then they will not serve as role-relative reasons for action distinct from those considered in the previous section. They will not, that is, serve as reasons to think that morality might in principle demand the punishment of the justified.

75. Cf. Margaret J. Radin, *Reconsidering the Rule of Law*, 69 B.U. L. REV. 781 (1989) (arguing that the rule of law is a contested concept because it reflects both instrumental values, such as enhancing liberty, and substantive values, such as the embodiment of fairness).

76. E.g., Michael S. Moore, *A Natural Law Theory of Interpretation*, 58 S. CAL. L. REV. 277, 314 (1985).

77. *Id.* at 313. "[I]n a constitutional democracy where there is a commitment to adjudicating disputes rather than having them decided by judicial fiat, the rule of law requires that the institutional role and responsibility of the judge be that of applying the law rather than making it." SARTORIUS, *supra* note 16, at 179.

1. *The Protection of Individual Liberty*

The first value that the rule of law is thought to serve is that of personal liberty. Liberty is enhanced when individuals can predict the consequences of their actions. To the extent that those consequences are variable (so as to make their prediction impossible), or secret (so as to preclude their discovery), or vague (so as to prevent their accurate assessment), individuals are handicapped in their ability to make plans or chart future conduct. Insofar as this handicap chills industry, it chills liberty.⁷⁸

Because judges alone wield the power of punishment, judges appear uniquely situated to affect the consequences of individual actions. For this reason, the value of liberty appears of concern to judges in a manner in which it is not of concern to private individuals. While individuals can affect the liberty of others by responding to their actions in unpredictable ways, they can seldom affect as many individuals in as substantial a manner as can judges who impose punishment in unforeseeable circumstances. Thus, while the citizen who contemplates the violation of the law must factor into her deliberations the likelihood that her illegal conduct will be unpredictable to others, and hence liberty-limiting, this reason for obedience appears unlikely to weigh as heavily in her calculations as it weighs in the calculations of a judge who contemplates the violation of the law for purposes of acquitting a justified offender. As Fuller maintained, because the law, unlike any other institution, affects the conduct of *all* citizens, liberty interests are uniquely jeopardized when the consequences of the law become unknown, unclear, variable, or contradictory.⁷⁹ If law is to protect and enhance liberty, its mandates must be clear and its penalties obvious.

If judges can make penalties obvious only by applying them to all offenders, unjustified and justified alike, then judges have a role-relative reason to punish the justified. We need not, however, pursue this defense of role-relative morality at any great length, for its success clearly depends on the assumption that citizens are prone to error. Its implicit presupposition is that citizens are often unlikely to see exceptions for what they are, and are thus likely to believe erroneously that a rule has been altogether altered when it is only the case that an exception to that rule has been made. They are thus likely to change

78. See SCHAUER, *supra* note 35, at 137-41. Schauer treats liberty-based arguments for judicial rule-following as arguments for fostering reliance. Because these are not synonymous, I find it useful to distinguish the arguments for liberty from the arguments for protecting reliance interests. See *infra* text accompanying notes 82-100.

79. See FULLER, *supra* note 15, at 38-41.

their course of conduct, or abandon it altogether, when their predictions are confused by the erroneous belief that the standards that apply to that conduct have been changed. They will attempt to conform their behavior to the example set by the justified offender in instances in which that behavior is unjustified. When penalized for violating what they mistakenly thought was an old rule, they are likely to be confused into paralysis. To prevent this paralysis, rules must be promulgated without the exceptions that are likely to be erroneously construed as contradictions. Insofar as the acquittal of justified offenders conveys contradictions to those prone to error, justified offenders should be punished.⁸⁰

If individuals were free from error, however, they would not mistake the acquittal of justified offenders for an indication that the rules to which they should conform their conduct have been altered. They would recognize that they would be entitled to violate the law in circumstances relevantly similar to those in which the justified offender violated it, but in no others. That is, the punishment of the justified would not serve as a necessary means of preserving the clarity of the law or the predictability of the penalties that attend its unjustified violation.

Because the protection of liberty might justify the punishment of the justified only in a world in which individuals were likely to misconstrue the significance of acquittals, a defense of role-relative morality that relies on the value of liberty collapses into an error argument. While judges seemingly have a reason to punish the justified if doing so prevents liberty-limiting errors by others, this reason should be counted among those that provide pragmatic rather than principled foundations for role-relative morality.⁸¹

80. This error argument is a version of the argument from example. See *supra* text accompanying notes 52-57. Its claim is that the judicial acquittal of a justifiably disobedient citizen will be erroneously believed to constitute a message that the law making the citizen's conduct an offense has been changed. Other citizens will then engage in such conduct indiscriminately, thus violating the still-existing law more often than is justifiable.

81. Moreover, if construed as a version of the error thesis, the argument from liberty may very well fail to establish a convincing case for role-relative morality. While our society highly prizes the protection of liberty through the rule of law, it has nevertheless adopted penal codes that embody a general balance-of-evils defense to what would otherwise constitute criminal violations. See MODEL PENAL CODE § 3.02 (Proposed Official Draft 1962), *supra* note 27. In Part II, I charted four sorts of cases in which justified offenders are unable to avail themselves of this legal defense. But a considerable number of cases remains in which offenders may escape punishment because the harm done by their offenses is less than the harm averted by those offenses. Insofar as we give legal recognition to the moral justification for such offenses, we clearly do not think that citizens are so prone to error as to construe all acquittals as acts that repeal rules against killing, stealing, destroying property, and so forth. At most, such acquittals will leave citizens in doubt about whether particular killings or particular acts of theft would be justified on their part under the balance-of-evils defense. We rightly regard this doubt as insufficiently liberty-limiting to warrant the wholesale punishment of justified offenders.

2. *The Protection of Reliance Interests*

The second rule-of-law value to which those who seek a principled foundation for role-relative morality might appeal is the value of protecting reliance. Three arguments can be distinguished for the importance of preserving reliance interests. The first is an argument from fairness. When individuals *justifiably* rely upon another's actions or statements and change their positions accordingly, fairness dictates that they should not be surprised by having their expectations thwarted and their positions changed for the worse. The second is an argument from industry. Even when individuals *unjustifiably* rely upon the statements or actions of another, their reliance should be protected as a means of ensuring that their efforts are not wasted. The third is an argument from the value of coordination. Where reliance is necessary for the solution of coordination problems and prisoner's dilemmas, that reliance should be protected as a means of achieving the collective goods that coordination makes possible.

a. *The argument from fairness.* Fairness demands that, if one encourages others to alter their positions, one should not leave those individuals worse off by failing to fulfill their expectations.⁸² If it is unfair to thwart the expectations of those who justifiably alter their positions in reliance on one's words or deeds, then, other things being equal, neither citizens nor officials should act contrary to the way in which they have encouraged others to expect them to act. Thus, the argument from fairness appears to provide both citizens and judges with a reason to act in accordance with the justified expectations of others. Insofar as citizens justifiably rely upon one another to abide by the law, citizens have reasons (in some instances, very weighty reasons) to obey laws that in the absence of any reliance interests would be justifiably disobeyed. Similarly, insofar as citizens and other officials justifiably rely on judges to abide by the law, judges have reasons (perhaps very weighty ones) to reach decisions that in the absence of such reliance interests might not be required.⁸³

The argument from fairness would give rise to differential reasons for obedience only if the justified expectations concerning citizens' conduct differed from the justified expectations concerning judges' de-

82. "[I]t is confessedly unjust to *break faith* with anyone — to violate an engagement, either express or implied, or disappoint expectations raised by our own conduct, at least if we have raised those expectations knowingly and voluntarily." MILL, *supra* note 5, at 285.

83. The argument from fairness constitutes the basis of what Lon Fuller called "the bond of reciprocity" between the citizen and the state. "Government says to the citizen in effect, 'These are the rules we expect you to follow. If you follow them, you have our assurance that they are the rules that will be applied to your conduct.'" FULLER, *supra* note 15, at 39-40; see SARTORIUS, *supra* note 16, at 166.

cisions. In particular, the argument from fairness would justify the punishment of a justified offender only if there existed legitimate expectations of the offender's punishment that differed from the legitimate expectations of the offender's obedience. For *eo ipso* a justified offender is one who rightly calculates that the reasons for obedience provided by the justified expectations of others are outweighed by the reasons for disobedience. For a judge to be justified in punishing such an offender on grounds of fairness to others, the judge's failure to punish would have to threaten sufficient unfair surprise to tip the balance of reasons for judicial action in favor of punishing an individual for whom unfair surprise did not tip the balance of reasons for private action in favor of obedience.

Four conditions appear necessary to a claim of unfair surprise: (1) one must in fact believe that another will do a certain action; (2) one's belief must be reasonable, that is, epistemically justified; (3) one must change one's position as a result of that belief in a manner that is potentially costly or detrimental to oneself; and (4) one's change of position must not itself be morally reprehensible. In order for justified reliance to provide a judge with a reason to punish a justified offender, some class of persons must meet the above four conditions. But who might *justifiably* rely on the punishment of the justified? Three categories of persons come to mind: the offender herself; those who prosecute the offender (either criminal prosecutors or private litigants); and members of the public at large. In order to assess whether any of these might justifiably rely on the punishment of a justified offender, one would need to work through the reasons to think that one or more of them might meet the four conditions specified above. There are several reasons to think that while offenders, prosecutors, and citizens-at-large may justifiably believe that justified offenders will be punished⁸⁴ they do not alter their position to their potential detriment as a result of that belief.⁸⁵ As such, they cannot be unfairly surprised by a refusal to punish the justified.

84. These individuals might in fact believe that justified offenders will be punished, and they might be justified in so believing because, as we explored earlier, the criminal law and civil law so provide. See *supra* notes 24-35 and accompanying text.

85. In cases of civil disobedience in which offenders seek punishment as a means of calling public attention to an immoral law, offenders might be thought to alter their positions in a potentially adverse manner. Consider the case of a war protester who avoids the draft just because his expected punishment is likely to cause public outrage and resentment over the war. If his disobedience is found to be justified and he is acquitted, he will no doubt have cause to complain that he is worse off than he would have been had he directed his activist energies toward other projects.

Such cases are relatively rare. Typically, while a justified offender might be surprised at her acquittal, she will not be unfairly surprised, any more than an unjustified offender will be unfairly surprised by mercy. Unfair surprises are unfair precisely because they make persons worse off. Because acquittals typically make offenders better off, they typically come as pleasant surprises.

Yet, even if offenders, prosecutors, or members of the public alter their positions to their potential detriment on the basis of a justified belief that judges will punish justified offenders, their reliance will not be justified unless it is morally legitimate. Individuals are unfairly surprised when their expectations are thwarted only if their expectations are not morally reprehensible. Thus, one is not justified in relying on another's performance of an immoral promise (to kill one's enemy, for example), even if one has every reason to believe that the promise will be kept, and even if one alters one's position substantially because of it. Insofar as the act relied upon is immoral, one's own reliance on that act is morally unjustified. Hence, one cannot complain of an unfair surprise when the would-be killer reneges on his promise.⁸⁶

Hence, in most cases, the offender herself will not have altered her position in a manner that will make her worse off if she is not punished.

Insofar as criminal prosecutors and civil plaintiffs justifiably believe that a justified offender will be punished or civilly sanctioned, their prosecution of such an offender will leave them worse off if the offender is not punished. While surprising public prosecutors may seem less unfair than inefficient, surprising private litigants smacks of substantial unfairness. If the law promises to sanction justified defendants, and if private litigants spend substantial amounts of time and money to bring suit against such defendants as a result of that promise, they will surely find themselves substantially worse off if that promise is breached by judges. Hence, the reliance interests of plaintiffs, and to a lesser degree of criminal prosecutors, might provide judges with a reason to impose civil sanctions or criminal punishment on justified offenders.

Members of the public who reasonably believe that justified offenders will be punished may rightly suppose that such punishment will deter the justified from disobeying the law. They may therefore fail to take precautions that they would otherwise take if they anticipated justified disobedience. They might coast through green lights without looking for drivers who might be justified in running opposing red lights — for they might rightly suppose that such drivers would be deterred from doing what they would be justified in doing.

Yet if this argument provides a reason for judges to think that members of the public might have changed their positions for the worse as a result of the expectation that justified offenders will be punished, it seemingly also provides a reason for would-be offenders to obey the law. That is, if members of the public expect that others will obey the law because they will be punished even if they disobey it justifiably, then those who contemplate disobedience must factor in this expectation as a reason to obey the law. If they do so accurately, and if the balance of reasons for action still favors disobedience, then it would seem that judges should not punish such offenders for thwarting public expectations, because the citizen's calculation of public reliance will exhaust a judge's calculation of that reliance. Were one to resist this by arguing that acquittals will reduce the degree to which justified offenders will take seriously the public's expectation that they will obey the law out of fear of punishment, one would smuggle in assumptions about error. One would suppose that justified offenders will come to make errors — but if they did, they would cease to be justified. Hence, while civilly disobedient offenders and public and private prosecutors may entertain reliance interests that are of unique concern to judges, members of the public probably do not.

86. Consider, as a second example, the reliance claim at stake in the famous case of *MacPherson v. Buick Motor Co.*, 111 N.E. 1050 (N.Y. 1916). Buick believed that it would not be held liable for injuries caused to consumers as a result of defects in the manufacturing of its automobiles. It justifiably believed this, because the law had upheld privity of contract in product liability cases, thus allowing auto manufacturers to be sued only by distributors (who were rarely injured by defects in the cars that they sold to consumers). Buick priced its cars on the assumption that it would not be held liable to remote purchasers, and so changed its position to its potential detriment. But Buick's use of the privity doctrine to escape responsibility to those injured by its negligence was morally reprehensible. Buick, after all, had both a moral and a legal duty to build safe automobiles, and the privity limitation on which it relied to escape liability was

Just as reliance on the performance of a promise will not constitute *justified* reliance if the promise is an immoral one, so reliance on the punishment of the justified will not constitute *justified* reliance if such punishment is itself unjustified. But whether it is justifiable to punish the justified is precisely the question to which reliance was supposed to provide an answer. Those who have supposed that there are role-relative reasons to punish the justified cannot, without circularity, maintain that justified reliance is one of them — at least not until they have found others that vindicate the morality of punishing the justified so that reliance on that punishment is in fact morally justified.

Thus, while an appeal to justified reliance is virtually a knee-jerk reaction for many lawyers and legal theorists who are called upon to advance a reason to think that judges should follow the rules, it is crucial to see that this reaction should be tempered by the recognition both that justified reliance may be difficult to make out and that justified reliance depends on arguments other than reliance to make it justified. If, however, there are arguments other than reliance that justify the punishment of the justified, then reliance on that punishment will itself be justified (so long as the other conditions for such reliance are satisfied). Thus, one need not conclude that justified reliance does not provide at least a *prima facie* reason to punish the justified — for it does provide such a reason if there are (also) other reasons to punish the justified.

b. The argument from industry. Those who would defend role-relative morality might circumvent the charge of circularity in the previous section by arguing that, even if individuals *unjustly* rely on the punishment of justified offenders, that reliance nevertheless provides a reason (though not one of fairness) for judges to administer such punishment: namely, unjustified reliance produces efforts that in some circumstances ought not to be wasted.

Consider a case that does not involve the issue of punishing the justified, but that clearly demonstrates the moral force of the argument from industry. In *Tennessee Valley Authority v. Hill*,⁸⁷ the Tennessee Valley Authority (TVA) received millions of dollars from the House Appropriations Committee to build the Tellico Dam. Upon discovering that the dam would render extinct a species of three-inch fish known as “snail darters,” the TVA consulted with the House Appropriations Committee to determine whether the project was barred by

only a limitation on the legal remedies available to consumers to redress breaches of Buick's duty. Hence, Buick was not *unfairly* surprised when Justice Cardozo concluded that privity would not bar a suit by a remote purchaser.

87. 437 U.S. 153 (1978).

the Endangered Species Act. Having received assurances from the Committee that the Act did not prevent construction of the dam, the TVA completed the project. In subsequent litigation, this reliance by the TVA on the Committee's statutory interpretation was deemed unjustified, for the Committee was neither a court charged with statutory construction nor a legislative body capable of modifying the terms of the Endangered Species Act. Nevertheless, it was thought that the industry and expense which such reliance engendered should not be casually wasted. Thus, the Supreme Court came in for heavy criticism when it subsequently enjoined the operation of the dam, the critics charging that the Court failed to attribute sufficient weight to the multimillion dollar waste that would occur as a result of its decision.

In the context of the present discussion, the argument from industry would appear to provide a role-relative reason for punishing the justified in circumstances in which individuals rely upon that punishment (however unjustifiably) and invest resources that will go to waste in the event that such punishment is not administered. Consider the following case of just such reliance. In *United States v. Kirby*,⁸⁸ a county sheriff was prosecuted under a federal statute that made it a crime to "obstruct or retard the passage of the mail, or of any driver or carrier"⁸⁹ after he carried out a warrant to arrest an on-duty federal mail carrier suspected of murder. The federal prosecutor in this case unjustifiably relied upon the Court to enforce the letter of the law over its spirit, and thus invested prosecutorial resources in the expectation that a justified offender would be punished.⁹⁰

Because those who would seek a source of role-relative morality need only establish that there are some reasons for action unique to judges, it might be enough to point to the waste of resources that accompany instances of unjustified reliance and argue that such a waste, while not a weighty reason for punishing the justified, is nevertheless some reason for punishing them.⁹¹ But if unjustified reliance is the sole source of role-relative morality, role-relative morality is sufficiently weak to cause us little concern. Even if unjustified reliance constitutes a role-relative reason for action, it is unlikely to justify the punishment of the justified, for it is unlikely to weigh more heavily

88. 74 U.S. (7 Wall.) 482 (1868).

89. 74 U.S. at 483 (quoting 4 Stat. 104 (1825)).

90. He was not. As the Court concluded: "All laws should receive a sensible construction. General terms should be so limited in their application as not to lead to injustice, oppression, or an absurd consequence." 74 U.S. at 486.

91. Such an argument might lead its proponents to the seemingly absurd conclusion that, other things being equal, a judge should decide in favor of the litigant who spent the most resources to litigate the case before the court.

than the value of acquitting those who do the right thing.⁹²

c. *The argument from coordination.* The protection of reliance interests is often thought important on grounds other than fairness or efficiency. Insofar as coordination solutions require individual reliance on the cooperation of others, the collective goods that they produce may depend upon the protection of such reliance interests. That is, if the advantages of coordination can be achieved only if citizens comply with coordinating conventions (like the rules of the road) and do not default on cooperative solutions to prisoner's dilemmas (such as those established by tax laws, environmental protection laws, and criminal laws), and if citizens will comply with such conventions only if they can rely on others's compliance with such conventions, then judges may have a reason to punish those who break laws that serve coordinating functions as a means of protecting important reliance interests.

But does this reason for punishment differ from any of those reasons that a citizen has to comply with the law in the first place? If not, then the argument from coordination cannot justify the punishment of the justified. *Eo ipso*, the justified offender has accurately assessed the importance of preserving the relevant coordination solutions and the degree to which her disobedience will affect the salience of those solutions. Her decision to break the law is justified precisely because she has given due weight to the value of others' reliance on her obedience and correctly calculated that that value is outweighed by the values accomplished by disobedience. For the judge to be justified in punishing the justified offender, the value of the reliance interests involved, or the importance of the coordinative schemes protected, must be greater for the judge than for the citizen.

There are tempting reasons to suppose that calculations concerning reliance interests do vary between citizens and judges. For a citizen calculates the reliance of others on *his* obedience, while a judge calculates the reliance of others on *her* obedience. Because the latter seems likely to be greater than the former, the argument from coordination seems to provide a plausible source of differential reasons for action.

i. *Prisoner's dilemmas.* Consider first the reliance interests at stake in cases involving prisoner's dilemmas. Prisoner's dilemmas arise in circumstances in which individuals would be best off if they could default from cooperative ventures without causing others to do

92. See Moore, *supra* note 76, at 386-88.

the same.⁹³ Thus, for instance, an individual would be best off if he alone could avoid paying taxes, union dues, and cable television pledges, for he would then be able to enjoy the public goods made possible through others' contributions without himself contributing to their support.⁹⁴ Because all individuals are similarly inclined, collective goods are fragile accomplishments. They can be sustained only if each individual can rely on (almost) all other individuals not to free ride on his cooperative efforts or contributions.

Those who defend role-relative morality point to the fact that individuals frequently appear fully justified in free riding on the cooperative efforts of others, notwithstanding the fact that such efforts depend on the absence of such free riding. That is, citizens appear justified in refusing to cooperate when they rightly believe that their refusal will not substantially affect the cooperative efforts of others or the collective goods that such efforts accomplished. This will be the case when the effect of an individual's contribution to a cooperative scheme is small, the impact of her default is negligible, and the gain from such a default is great. When union dues would be better spent on individual acts of charity, cable television pledges would be better spent on food and clothing for one's children, and the time devoted to voting would be better spent on virtually anything, individuals are justified in failing to pay fees or vote, so long as they rightly calculate that their individual failures will not substantially contribute to the dissolution of the union, the bankruptcy of the cable company, or the collapse of democracy.⁹⁵

Many believe that the primary purpose of law is to provide cooperative strategies by which to achieve collective goods and, when necessary, to induce compliance with those strategies by means of sanctions.⁹⁶ Such sanctions accomplish two ends. First, the threat of

93. See GAUTHIER, *supra* note 7, at 79-82; PETER SINGER, THE EXPANDING CIRCLE 45-48 (1981); EDNA ULLMANN-MARGALIT, THE EMERGENCE OF NORMS 25-26 (1977); David P. Gauthier, *Morality and Advantage*, in MORALITY AND RATIONAL SELF-INTEREST, *supra* note 7, at 166, 170-71.

94. Prisoner's dilemmas can function positively in circumstances in which cooperation is dangerous. For an illuminating discussion of the advantages of noncooperation in the area of antitrust law, see ULLMANN-MARGALIT, *supra* note 93, at 44-45.

95. [I]t is a melodramatic exaggeration to suppose that every breach of law endangers, by however small a degree, the survival of the government, or of law and order. Many acts of trespass, breaches of contract, violations of copyright, and so on, regrettable as some of them may be on other grounds, have no implications one way or another for the stability of the government and the law.

RAZ, MORALITY, *supra* note 36, at 102.

96. This is the argument to which John Finnis has devoted his entire book, *Natural Law and Natural Rights*. See FINNIS, *supra* note 11. In advancing the claim that the essential function of law is a coordinative one, Finnis follows St. Thomas Aquinas. See THOMAS AQUINAS, *The Summa Theologica*, in 2 BASIC WRITINGS OF SAINT THOMAS AQUINAS 225, 744-45, 750-52

punishment provides individuals with a new reason for action that ordinarily tips the balance of reasons for action in favor of cooperation.⁹⁷ Second, this guarantees that the individual can rely on the fact that she will not be one of the sole contributors to a scheme predominately enjoyed by free riders. The ability to rely on the cooperation of others thus makes the individual's own cooperation rational.

Were judges to refuse to punish justified offenders, they would destroy the incentives that make collective goods possible. Individuals would cease to calculate the pain of punishment among the reasons for compliance; absent that factor, the rational course of conduct would be a noncooperative one. Hence, judges have weighty reasons to punish those who, absent punishment, would have weighty reasons to refuse to follow the law in situations that constitute prisoner's dilemmas. That is, the argument from coordination appears to justify the punishment of justified offenders in cases in which that punishment is required to solve a prisoner's dilemma.

ii. Coordination problems. Now consider the reliance interests at stake in cases involving coordination problems rather than prisoner's dilemmas. Coordination problems arise when individuals seek to cooperate (rather than to free ride on the cooperation of others) but are unable to settle on a means of doing so.⁹⁸ Classic examples include

(Anton C. Pegis ed., 1945); see also John M. Finnis, *Law as Co-ordination*, 2 *RATIO JURIS* 97 (1989); Chaim Gans, *The Normativity of Law and Its Co-ordinative Function*, 16 *ISRAEL L. REV.* 333 (1981); Leslie Green, *Law, Co-ordination and the Common Good*, 3 *OXFORD J. LEGAL STUD.* 299 (1983).

97. RAZ, *MORALITY*, *supra* note 36, at 50-51; SINGER, *supra* note 93, at 46; Hurd, *supra* note 43, at 1019-21.

There is a large and well-known literature devoted to vindicating the claim that morality itself provides sufficient reason for cooperation in prisoner's dilemma situations. See, e.g., GAUTHIER, *supra* note 7, at 8-10, 113-56; SINGER, *supra* note 93, at 47.

98. For a deeper appreciation of the nature of coordination problems, see DAVID K. LEWIS, *CONVENTION: A PHILOSOPHICAL STUDY* (1969); THOMAS C. SCHELLING, *MICROMOTIVES AND MACROBEHAVIOR* (1978); THOMAS C. SCHELLING, *THE STRATEGY OF CONFLICT* 89-99 (1960); ULLMANN-MARGALIT, *supra* note 93, at 18-73. For discussions of coordination problems that are more specifically jurisprudential, see FINNIS, *supra* note 11, at 231-59; RAZ, *supra* note 3, at 64, 159; RAZ, *MORALITY*, *supra* note 36, at 49-50; Lon L. Fuller, *Human Interaction and the Law*, 14 *AM. J. JURISPRUDENCE* 1 (1969); Gans, *supra* note 96; Green, *supra* note 96; Hurd, *supra* note 43, at 1016-19; Gerald J. Postema, *Coordination and Convention at the Foundations of Law*, 11 *J. LEGAL STUD.* 165 (1982); Joseph Raz, *Authority and Consent*, 67 *VA. L. REV.* 103 (1981).

On most accounts, coordination problems are not the product of human error. Even individuals who possess superb practical rationality may find themselves in circumstances that demand coordination with others but provide no salient means by which to coordinate. This is because coordination problems result from a plurality of *equally* moral and *equally* practicable coordinative solutions. As Aristotle put it, a natural rule of justice is one "which everywhere has the same force and does not exist by people's thinking this or that . . ." A conventional or "legal rule" is a rule "which is originally indifferent, but when it has been laid down is not indifferent. . . ." ARISTOTLE, *supra* note 1, at 1790-91. Only if one were prepared to defend a rigorous right-answer thesis — a thesis that held that morality provided a right answer to every normative question (including whether persons should drive on the right or the left, shake hands with their

instances in which individuals share a desire to avoid a collision but lack a convention that will coordinate their actions so as to eliminate the risk of that collision.

The solution to a coordination problem depends upon widespread recognition of a single, salient cooperative strategy.⁹⁹ While the law provides such strategies in many circumstances that demand coordination, the salience of those strategies may depend in large part on their judicial enforcement.¹⁰⁰ This is true for two reasons. First, judges are generally a salient source of information about the law. Hence, citizens who seek to coordinate their efforts with others will look to judges for information concerning the coordination solutions that have been provided by law, for they justifiably expect that others will do likewise. Second, because the law will provide a solution to a particular coordination problem only if people in fact abide by the law (as opposed to some other coordinating convention), and people are more likely to abide by the law if they are punished for noncompliance, citizens will look to judges for evidence that they are punishing noncom-

right or their left hand, place their forks on the right or the left, etc.) — could one maintain that coordination problems result from persons' erroneously failing to see the singularly right course of conduct that morality prescribes in those circumstances.

99. David Lewis explains:

Some combinations of the agents' chosen actions are *equilibria*: combinations in which each agent has done as well as he can given the actions of the other agents. In an equilibrium combination, no one agent could have produced an outcome more to his liking by acting differently, unless some of the others' actions also had been different.

LEWIS, *supra* note 98, at 8. Gerald Postema elaborates:

Solutions to coordination problems are based on each party's exploiting mutually concordant expectations. . . . Since what *I* do depends on what *you* will do, in the ideal case I attempt to replicate your practical reasoning to determine what you will do. And since I know that what you want to do depends on what I do, I must, in replicating your reasoning, determine what you expect *me* to do. And since you are engaged in the same process with regard to me, to replicate your reasoning I must replicate your attempt to replicate mine, and so forth. Given this framework for the nesting of expectations, all that is needed to break the deadlock of a coordination problem is some fact about one of the equilibria which isolates it from the others and which is obvious to both of us and known by us both to be obvious to the other. Thus successful coordination requires the parties to locate some *salient* fact about one of the equilibria that makes it stand out, that is, to read the same message in the common situation, and with that message converge on a solution.

Postema, *supra* note 98, at 174; see also Green, *supra* note 96, at 301-02 (discussing development of coordinative equilibria).

100. Paul Weiler has advanced the argument that judicial rule-following is also crucial to fostering the sort of coordination necessary for private settlement of legal disputes.

Since the vast majority of tort actions, criminal prosecutions, and so on are privately negotiated and settled, and the continuance of this practice is absolutely essential to staving off the breakdown of the judicial system, we cannot afford to do anything which lessens the incidence of such private settlements. The theory is that the best way to ration the costly process of adjudication is to allow private individuals to decide rationally that the likely gains are outweighed by the likely costs. Only if there is a substantially accurate awareness of the standards the courts will use will such prediction and negotiation be rational.

Paul Weiler, *Legal Values and Judicial Decision Making*, 1970 CAN. BAR REV. 11 (1970).

pliance, for the fact of such punishment confirms the law's role as the dominant or salient coordination convention.

Of course, citizens can themselves affect the salience of a particular coordinative strategy. Their failure to comply with a particular coordination solution will make that solution less salient, for it will make it the case that others cannot as readily rely on it as a solution, and thus, it will provide an impetus to seek other bases of coordination. But there are two reasons to suspect that individual departures from legally established coordination conventions will not affect the salience of those conventions as dramatically as will the judicial refusal to enforce them. First, individual departures are unlikely to achieve widespread recognition, and hence, they are unlikely to shake the faith of the majority in the salience of a particular coordinative strategy. Second, coordination solutions, unlike solutions to prisoner's dilemmas, do not provide incentives to default. If most people drive on the right side of the road, it is in the individual's (self-) interest to do so as well. Knowledge of the fact that coordination solutions are in everyone's best interests provides individuals with a reason to interpret an actor's noncooperation as a product of ignorance, irrationality, or emergency. Unless they have a reason to suspect that a substantial number of others will alter their course of conduct by virtue of the (ignorant, irrational or imperiled) actor's noncompliance, they have no reason to think that that actor's noncooperation makes a widely recognized cooperative strategy less salient.

There thus appear reasons to suppose that the reliance interests at stake in situations that pose coordination problems vary between citizens and judges. A citizen may rightly calculate that running a red light in the middle of the night will not substantially affect the practice of stopping for red lights, but a judge may rightly calculate that her acquittal of that citizen will quite substantially affect that practice. Insofar as citizens look to the judicial enforcement of the rules of the road as evidence that others will abide by those rules, an acquittal may cause citizens to think that others will alter their behavior. And this by itself will make it rational for them to alter their behavior accordingly. An acquittal may thus have a domino effect that forces citizens to look beyond the law for alternative coordination schemes by which to reduce traffic injuries at intersections. Insofar as the failure of a judge to punish an offender is likely to affect the salience of a coordination solution upon which others depend, while the failure of the citizen to abide by such a solution is not, a judge will have a more weighty reason to abide by the decision rule requiring the punishment of a disobedient citizen than that citizen will have to abide by the conduct

rule that assures coordination. The argument from coordination thus appears to support role-relative morality.

We should, however, be hesitant about this conclusion. For the argument from coordination threatens to collapse into the argument from example by smuggling in assumptions about the propensity for error on the part of citizens. That is, the argument seems to suggest that a judge should punish justified offenders so as to affirm the salience of a convention for those who might mistakenly take an acquittal to be grounds for thinking that a particular convention should no longer be trusted. Morally perfect reasoners, on the other hand, could be expected to see an exception to a conventional rule for what it is, namely, an exception. Such an exception ought not to shake their faith in the salience of a particular coordination rule, so long as they have no reason to suspect that that exception will become the rule. Hence, in a world devoid of error judges could afford to recognize such exceptions by acquitting justified offenders without fear that they would promote unjustified departures from genuinely salient coordination solutions.

But herein lies the rub. Recognition of exceptions to coordination solutions only creates new coordination problems that demand solutions. Even if citizens, by virtue of being capable moral reasoners, were not induced to abandon coordinative strategies by virtue of the judicial recognition of justified departures from those strategies, they would nevertheless require means by which to coordinate with justified offenders. And if there were second-order conventions that allowed citizens to coordinate their conduct with justified offenders of first-order conventions, judges would have to recognize exceptions to those second-order conventions in cases in which individuals justifiably departed from those second-order conventions. Such exceptions would create, once again, coordination problems for those seeking to coordinate their conduct with second-order justified offenders, thus necessitating third-order conventions. The potential for regress looms large.

To appreciate the regress, consider the following. Traffic lights currently provide first-order coordination solutions for drivers who meet one another at intersections. There are, however, recognized instances in which individuals are justified in violating traffic signals. Ambulance drivers, police officers, and firefighters often have reasons to run red lights that exceed the reasons to stop at them. Recognizing this fact, and recognizing that such a fact creates a new coordination problem for those who seek to coordinate their actions with the justified violations of such officials, a second-order coordination solution

has been created by the use of flashing lights and sirens. Such devices signal a violation of the first-order convention and hence solve the second-order coordination problem that recognition of such potential violations creates.

Now imagine the following case. An ambulance transporting an individual who is injured but in stable condition approaches an intersection with lights flashing and sirens blaring. An unmarked private car transporting an individual who is near death approaches the same intersection at the same speed. Both drivers confront red lights. If they simultaneously run their red lights, they will collide with each other. And because the private car is unmarked, it may well be hit by those who are seeking to comply with the second-order convention that requires them to yield to the ambulance. (This admittedly takes a complex intersection, so we should suppose that this takes place in New Jersey.) The driver of the car rightly supposes that he has reason to violate the first-order convention that requires him to stop at the red light. He also rightly guesses that it is more important for him to get his patient to the hospital before the ambulance driver gets her patient to the hospital. He thus rightly supposes that he has reason to violate the second-order convention that requires him to yield to the ambulance. He takes seriously the fact that there is no convention that governs this situation, and that neither the ambulance driver nor other private drivers will recognize the justifiability of this violation so as to yield to it. He nevertheless rightly calculates that the life at stake justifies the risk of a collision.

Were a judge to acquit the driver of the car and thereby carve out an explicit exception to the second-order convention that now governs the justified violation of first-order traffic conventions, the judge would create the need for a third-order convention that would enable ambulance drivers and other citizens to yield to individuals who justifiably violate both the first-order and second-order conventions. And this convention would plainly have its justified violations, which, if judicially recognized, would lead to the need for a fourth-order convention, and so on, *ad infinitum*.

Those who defend role-relative morality would be justified, *prima facie*, in supposing that this regress could be curbed (albeit not wholly eliminated) by the punishment of justified offenders. The punishment of justified offenders would provide such offenders with a new and weighty reason to comply with recognized first-order (or in rare cases, second-order) conventions. Knowledge of this fact by others would return their confidence in the general salience of the first-order conventions that make coordination possible.

If the acquittal of justified offenders generates a regress that defeats coordination altogether, then judges seemingly have a reason to punish justified offenders. For coordination is necessary in order to achieve certain collective goods, and such coordination is only possible if judicial decrees provide salient courses of conduct. If judicial decrees are susceptible to a recursive set of exceptions, then they will not provide salient courses of conduct. Hence, the judicial promulgation of a coordination strategy must be exceptionless for it to do its task, and it is seemingly exceptionless only if justified offenders are punished.

3. *The Protection of Equality*

The third and final rule-of-law value to which those seeking a principled foundation of role-relative morality might turn is that of equality. Equality requires the similar treatment of those who are identical in morally relevant respects. Conversely, it permits (and perhaps requires) the differential treatment of those who are not identical in morally relevant respects.¹⁰¹

There exists a long-standing controversy over the question of whether equality functions as an independent value at all,¹⁰² and I do not propose to add footnote fodder to that debate. Suffice it to say that, if equality does function as a genuine value, then it provides a reason to treat present cases like past cases, even when those past cases were dealt with unjustly or erroneously. Similarly, it provides a rea-

101. For the classic statements of these requirements of the principle of equality, see ARISTOTLE, *supra* note 1, at 1785-86; ARISTOTLE, *The Politics*, in 2 THE COMPLETE WORKS OF ARISTOTLE, *supra* note 1, at 2031-33, 2035-36 (Benjamin Jowett, trans.); PLATO, *Laws*, in THE COLLECTED DIALOGUES OF PLATO 1225, 1336-37 (A.E. Taylor trans., Edith Hamilton & Huntington Cairns eds., 1961); PLATO, *The Republic*, in THE COLLECTED DIALOGUES OF PLATO, *supra*, at 575, 786. See generally PETER WESTEN, SPEAKING OF EQUALITY 32-33, 181-229 (1990).

102. Some have argued that the principle of equality is, or can be, grossly unjust. See Larry Alexander, *Constrained by Precedent*, 63 S. CAL. L. REV. 1 (1989); William K. Frankena, *The Concept of Social Justice*, in SOCIAL JUSTICE 1, 17 (Richard B. Brandt ed., 1962); Phillip Montague, *Comparative and Noncomparative Justice*, 30 PHIL. Q. 131, 133 (1980). Others have argued that it is merely empty. Its requirements "possess both more truth and less content than is sometimes supposed: more truth because they not only happen to be true but are necessarily true; less content because, being necessarily true, they add nothing to what we already know." WESTEN, *supra* note 101, at 186 (referring to Aristotle's statements regarding equality); see KENNETH CAUTHEN, THE PASSION FOR EQUALITY 5 (1987); HANS KELSEN, GENERAL THEORY OF LAW AND STATE 439 (1945); W. VON LEYDEN, ARISTOTLE ON EQUALITY AND JUSTICE: HIS POLITICAL ARGUMENT 5 (1985); Peter Westen, *The Empty Idea of Equality*, 95 HARV. L. REV. 537 (1982).

Fred Schauer has convincingly advanced the surprising argument that, if equality is a value, rule-based decisionmaking jeopardizes rather than protects it. "When rule-based decision-making prevails, what increases is the incidence of cases in which relevantly different cases are treated similarly, and not the incidence of cases in which like cases are treated alike." SCHAUER, *supra* note 35, at 137.

son to treat present cases like future cases if one can predict the treatment that will be administered in those future cases, even if that treatment is now, and will be then, unjust or ill-conceived.

To the extent that equality is a value at all, it does not seem to be a value that is of exclusive concern to those occupying judicial roles. If parents let their son drive the family car on his sixteenth birthday, the value of equality speaks in favor of allowing their daughter to drive the car when she turns sixteen, even if they have realized, in the interim, that the decision in the son's case was an unwise one. Moreover, the value of protecting equality may properly enter into a citizen's practical deliberations concerning whether to break the law. If a bartender knows that other bartenders will follow the law that prohibits them from serving alcohol after 2 a.m., then the value of equality speaks in favor of refusing to serve alcohol after that hour, for doing so would give an unequal advantage to that bartender's small set of patrons.

Notwithstanding the fact that equality concerns may enter into the balance of reasons that determines what is right for a citizen to do, many suppose that their weight is greater when they enter into the balance of reasons that determines what is right for a judge to do. Thus, for example, many find intuitive the claim that a judge should seek comparative rather than substantive proportionality in sentencing,¹⁰³ even if the judge considers the severity of the established punishment substantively disproportionate to the severity of the offense.

The demand for comparative proportionality, or numerically equal punishment in like cases, creates a complex coordination problem for judges. If giving the same sentence that is given by others is more important than giving a sentence considered substantively proportional, then judges need a means of coordinating their sentencing decisions. Both common law precedents and legislatively enacted decision rules fit the bill, for they provide judges with salient means of ensuring

103. I follow both Plato and Aristotle in defining substantive proportionality as punishment proportional to desert, and comparative proportionality as numerically identical punishment. See ARISTOTLE, *supra* note 101, at 2067; ARISTOTLE, *Eudemian Ethics*, in 2 THE COMPLETE WORKS OF ARISTOTLE, *supra* note 1, at 1922, 1967-68 (J. Solomon, trans.); PLATO, *Laws*, *supra* note 101, at 1337. While Plato insisted that substantive proportionality was "the very award of Zeus," PLATO, *Laws*, *supra* note 101, at 1337, insofar as it is necessarily just, while comparative proportionality is only contingently just, Aristotle insisted that rectificatory justice is constituted by comparative proportionality. See ARISTOTLE, *supra* note 1, at 1785.

Modern sympathy for Aristotle's position often derives from the recognition that substantive proportionality is, at best, difficult to assess, and, at worst, arbitrary. Does mail fraud merit three years of imprisonment? Seven years? Nine years? Is two weeks of jail time proportional to the offense of indecent exposure, or two years? Perhaps the most that judges should aspire to is comparative proportionality, at least in cases in which the punishment that is comparative does not depart from the seemingly wide range of possible proportionately just punishments.

For a good discussion of the distinction between equality and proportionality, see WESTEN, *supra* note 101, at 5-7.

the equal treatment of litigants. Insofar as judges can be confident that other judges follow the law, they can be confident that the law provides them with accurate information about how cases similar to their own were adjudicated in the past and how they will be adjudicated in the future. It thus provides them with a reason to think that their own obedience will ensure the equal treatment of like cases.

If the value of protecting equality provides a weighty reason for judges to achieve comparative proportionality in punishment, and if comparative proportionality can be accomplished only by faithful adherence to the law, then judges have a reason to abide by the law that citizens do not. But does this fact provide a role-relative reason to punish the justified? After all, insofar as justified offenders are not morally culpable, they are morally dissimilar to unjustified offenders. Hence, concerns for equality would appear to favor dissimilar treatment.

Yet, if justified offenders were not treated differently in the past than were unjustified offenders or are not likely to be treated differently in the future, then a judge has a reason to impose comparable punishments on justified and unjustified offenders alike. Insofar as judges recognize that they can collectively accomplish comparative proportionality only if they coordinate their individual actions through obedience to the law, and insofar as the law instructs them to punish *all* offenders who satisfy the conditions of *legal* culpability and who are not justified or excused on *legally recognized* grounds,¹⁰⁴ judges have evidence both that other judges probably punished justified and unjustified offenders comparably in the past and that other judges will probably punish justified and unjustified offenders comparably in the future. The value of equality will thus seemingly provide a role-relative moral reason for a judge to punish justified offenders.

It is important to recognize that this argument does not smuggle in presuppositions about the need to protect against errors, and hence it functions, at least *prima facie*, as a principled argument for role-relative morality. It rests on the claim that, even if judges were perfect reasoners, they would still require a salient source of coordination, for morality makes equal punishment more important than substantively correct punishment. Once one judge expects others to follow the law so as to accomplish the equal treatment of like cases, and thus employs the law as a means of determining the appropriate punishment in a particular case, all other judges have a reason to do likewise. While the law's failure to distinguish the punishment due to justified and un-

104. See *supra* notes 24-35 and accompanying text.

justified offenders may itself be in error, the first judge's compliance with that law may not be erroneous if the law provides the *only* salient source of coordination and if coordination is indeed more important than substantive justice. Thereafter, the justifiability of punishing the justified increases as the number of cases in which justified offenders are punished multiplies.

Let us now pause to take stock of the argument thus far. Our task in this Part has been to determine whether there are at least *prima facie* reasons to embrace perspectivalism and abandon the punishment principle. If there are, then we have reason to think that the dilemma with which we began is solved, for morality will itself demand that we punish the justified so as to preserve morally valuable systemic commitments.

We began our analysis by canvassing the bases upon which concerns about moral error might provide judges with reasons to punish justified private offenses and institution designers with reasons to punish justified judicial offenses. We saw that error might provide a pragmatic reason to punish the justified but that it cannot provide a principled reason to do so. We then turned to the question of whether there are principled reasons for punishing the justified that are uniquely applicable to those within judicial roles. We saw that the protection of reliance interests and the preservation of equality appear to serve as more weighty concerns for judges than for citizens and therefore that, in some circumstances, the balance of reasons for judicial action might justify the punishment of a citizen for whom the balance of reasons justified disobedience. We now turn to the question of whether there are principled reasons for punishing the justified that are uniquely applicable to those who design and preserve legal institutions (such as the role of the judiciary). If such reasons exist, then, even if judges are justified in refusing to punish justifiably disobedient citizens, institution designers may be justified in punishing judges for such refusals.

C. The Values of Democracy and the Separation of Powers: Principled Foundations for Constitutional Perspectivalism

Fred Schauer has maintained that the standpoint of the designer of a decisionmaking environment is quite different from the standpoint of a decisionmaker within that environment.¹⁰⁵ If this is the case, then we must distinguish judicial perspectivalism from what I shall call

105. Frederick Schauer, *Rules and the Rule of Law*, 14 HARV. J.L. & PUB. POLY. 645, 691 (1991).

constitutional perspectivalism. Constitutional perspectivalism is the thesis that those who design and protect the decisionmaking institutions of the state — including judicial institutions — occupy a role that is characterized by reasons for action inapplicable to those in other roles.

The constitutional role is a nebulous one, because unlike the judicial role it no longer has a distinct set of occupants. In the absence of the Framers of the Constitution, the task has fallen to those who periodically step out of their other roles for the purposes of creating, evaluating, or policing decisionmaking systems that are essentially self-executing. Thus, the constitutional perspective is sometimes occupied by legislators, sometimes by appellate court judges and justices, sometimes by political activists, and sometimes by lawyers, legal scholars, and philosophers. But even if no one ever assumed the constitutional perspective, it would remain of philosophical interest. Insofar as such a perspective is available, and insofar as it generates reasons to discipline or impeach individuals who do the right thing from a different perspective, it raises the dilemma with which we began. The punishment principle is jeopardized in theory, even if there is no one to abandon it in practice. If its theoretical rejection is based on principled rather than pragmatic grounds, it must be admitted that morality is inherently paradoxical, for it licenses the punishment of those who act morally.

The task of this section is to explore the viability of constitutional perspectivalism. We shall seek *prima facie* reasons to think that those who do or could assume the constitutional perspective might be justified in punishing judges who are justified in acquitting justified offenders.

1. *The Classic Arguments for Structural Pluralism*

The discussion in Part II and the arguments advanced in the previous two sections of this Part combine to suggest that, while judicial obedience to the law may serve important values, there may nevertheless be circumstances in which judges should set aside the law and act in accord with their own best judgments. The recognition that judges may be compelled by the demands of practical reason to substitute their own judgments for those of the legislature flies in the face of our understanding of the principle of democracy and its concomitant demand for the separation of powers. For that principle has traditionally been thought to prohibit judges from setting aside legislatively enacted rules in the name of background moral considerations. Gov-

ernment by the people gives way to dictatorship by an elite if unelected judges can rewrite democratic legislation.

There is a vast literature devoted to defending structural pluralism. For our purposes, it is useful to think of this literature as embodying two alternative theories of the importance of democracy.¹⁰⁶ On one theory, democratic decisionmaking constitutes our most reliable means of achieving right results. As between competing sources of rules, democratic institutions are more likely to achieve rules that square with the balance of reasons for governmental action than are other institutions. Democratic institutions should therefore be accorded the power to make rules, while other institutions should be restricted to the lesser tasks of interpreting, implementing, and enforcing those rules. I shall call this first sort of theory an *instrumentalist* one, for its claim is that we should value democracy only to the extent that it achieves right results. In the event that democratic results fail to cohere with the balance of reasons for action, they lack any value at all. On this theory, we value democracy only because we value truth.

On the second sort of theory, democratic decisionmaking is intrinsically good. While it may fail to accomplish results that accord with the balance of reasons for action, it nonetheless instantiates certain values that other decisionmaking procedures do not. Insofar as living by wrong rules that reflect these values is morally preferable to living by right rules that do not, democracy is valuable even when it produces wrong rules. Hence, undemocratic institutions (like the judiciary) should not set aside the decisions of democratic institutions even when it is apparent that these decisions fail to reflect the balance of reasons for legislative action. To do so thwarts more important values than truth. I shall call this sort of theory an *internalist* theory, for its claim is that democratic decisionmaking is internally or inherently valuable.

In what follows, I shall canvass some of the classic instrumentalist and internalist arguments that have been made on behalf of democracy and the separation of powers. I shall demonstrate that each of these arguments provides a reason to punish disobedient judges only insofar as it provides a reason to think that judges who depart from legislatively enacted rules also depart from the balance of reasons for judicial action. That is, I shall show that each argument furnishes a basis for supposing that disobedient judges are *unjustified* offenders, and hence deserving of punishment. But I shall further demonstrate that, be-

106. I follow John Arthur's taxonomy of "instrumental" and "internal" theories of democracy. DEMOCRACY: THEORY AND PRACTICE xi (John Arthur ed., 1992) [hereinafter DEMOCRACY].

cause none of the classic arguments provides an exclusionary reason for judicial deference to the legislature, none of the arguments can exclude the possibility that judges might be *justified* in disobeying legislative rules. The most that they can provide is an additional reason for judges to comply with legislative rules. In the event that judges rightly conclude that the reasons for obedience furnished by these classic arguments are outweighed by reasons for disobedience, these arguments will license disobedience.

If I am right, these classic theories of democracy and the separation of powers cannot, themselves, justify the punishment of the justified. But they might well provide reasons to think that the argument from error, considered in section A of this Part, is applicable. If the classic arguments for democracy and the separation of powers provide reasons to think that judicial disobedience is more often in error than not, then they provide grounds for thinking that the acquittal of justified judicial offenders may trigger more unjustified judicial disobedience than justified judicial disobedience. These arguments might therefore ground an argument from error that would provide those who assume the constitutional perspective with a role-relative reason to punish judges who are justified in acquitting justifiably disobedient citizens. But as I have already argued, such a reason would constitute only a pragmatic basis for punishing justified judicial offenders. It will remain to be established, at the end of this section, whether any principled reasons remain to think that system designers might be justified in punishing justifiably disobedient judges.

a. Instrumentalist theories of democracy. According to the instrumentalist, the reasons for legislative action exhaust the reasons for judicial action. That is, the fact that the legislature reached a particular decision is not itself a reason for judicial compliance with that decision; it is not itself something valuable that must be added to the balance of reasons for judicial obedience. Judges should obey legislative enactments only because those enactments are more likely to conform to an antecedently existing balance of reasons for governmental action than are their own judgments.¹⁰⁷

i. The argument from metaethical relativism. John Ely has insisted that courts are incapable of deciphering better answers to social controversies than are legislatures, because there is no source for such answers beyond that provided by the results of democratic legislation. "[O]ur society does not, rightly does not, accept the notion of a discov-

107. Instrumentalist theories of democracy thus accord legislation only theoretical authority. See *supra* note 43 and accompanying text.

erable and objectively valid set of moral principles, at least not a set that could plausibly serve to overturn the decisions of our elected representatives."¹⁰⁸ In a similar vein, Robert Bork has argued that, when judges are called upon to decide cases according to moral principles, they have no means of deciding such cases "other than by reference to some system of moral or ethical values that has no objective or intrinsic validity of its own and about which men can and do differ."¹⁰⁹ Thus, the only principled means by which to resolve disputes is by reference to a source other than a judge's moral principles. When the Constitution is silent, judges act morally only if they conform their decisions to the will of the majority as expressed in democratically enacted legislation.¹¹⁰

At the heart of Ely's and Bork's claims is the metaethical thesis that morality is relative to the beliefs of a majority.¹¹¹ That is, what constitutes morality is whatever the majority of individuals within a community believes moral.¹¹² From this thesis, Ely and Bork take themselves to be in a position to argue that democratic outcomes constitute what is morally true about how persons should act. Any judge who arrives at a view contrary to the legislature's is necessarily in error. And nothing short of blind deference to the legislature can prevent such error.

Even if metaethical conventionalism is defensible, however,¹¹³ it does not support the claim that democratic results are *constitutive* of what is moral. Because a group of representatives may fail to capture the majority's beliefs when they vote for legislation, legislative results may depart from what the majority would in fact prefer. Legislation

108. JOHN H. ELY, *DEMOCRACY AND DISTRUST* 54 (1980); see Hurd, *supra* note 23, at 1446 n.121.

109. Robert H. Bork, *Neutral Principles and Some First Amendment Problems*, 47 IND. L.J. 1, 10 (1971).

110. *Id.* at 11.

111. See Hurd, *supra* note 23, at 1445-51.

112. See ROBERT L. ARRINGTON, *RATIONALISM, REALISM, AND RELATIVISM: PERSPECTIVES IN CONTEMPORARY MORAL EPISTEMOLOGY* 192-99 (1989); WALTER T. STACE, *THE CONCEPT OF MORALS* (1937); EDWARD WESTERMARCK, *ETHICAL RELATIVITY* (1932); BERNARD WILLIAMS, *MORALITY: AN INTRODUCTION TO ETHICS* 20 (1972); DAVID B. WONG, *MORAL RELATIVITY* 23 (1984); Philippa Foot, *Moral Relativism*, in *RELATIVISM: COGNITIVE AND MORAL* 152, 161 (Jack W. Meiland & Michael Krausz eds., 1982); Gilbert Harman, *Relativistic Ethics: Morality as Politics*, in 3 *MIDWEST STUDIES IN PHILOSOPHY* 109, 120 (Peter A. French et al. eds., 1980); Gilbert Harman, *Moral Relativism Defended*, 84 *PHIL. REV.* 3, 4 (1975); John Ladd, *The Issue of Relativism*, in *ETHICAL RELATIVISM* 107, 124-26 (John Ladd ed., 1973); Bernard Williams, *The Truth in Relativism*, 75 *PROCEEDINGS OF ARISTOTELIAN SOC.* 215, 215 (1974-75).

113. I have argued elsewhere that the beliefs or preferences of a majority cannot constitute moral facts concerning the rightness or wrongness of particular actions. If I am right, then the beliefs of the majority concerning how judges ought to decide particular cases cannot constitute moral facts that provide reasons to so decide cases. Hurd, *supra* note 23, at 1459-506.

thus cannot constitute majority sentiment; rather, it can only reflect it. As such, legislation can be, at most, good evidence of what the conventionalist considers moral. Accordingly, even the conventionalist must admit that it is possible for a judge to be right in concluding that morality demands a departure from legislative decisions (for a judge may accurately assess majority opinion while a legislature may not).¹¹⁴

While the truth of metaethical relativism would fail to establish that judicial departures from legislation are necessarily erroneous, it would nevertheless vindicate the claim that legislation is practically conclusive evidence of what is moral. It would therefore provide a reason to think that judicial departures from legislative rules are probably more frequently wrong than right. Yet, if a judge discovers a genuine discrepancy between democratic results and majority sentiment, the argument provides no basis for insisting that the judge ought nevertheless to abide by the democratic results. Hence, the argument provides neither an exclusionary reason for judges to comply with democratic results nor a basis for punishing the judge when she complies with majority sentiment rather than with democratic legislation. It provides, at most, a reason to think that judicial disobedience results more frequently from judicial error than judicial accuracy and thus that institution designers might be justified in punishing *all* acts of judicial disobedience as a means of reducing a disproportionate amount of unjustified judicial disobedience. But this argument for punishment is a version of the argument from error discussed in section III.A, and it thus provides only a pragmatic reason to punish justified judicial offenders.

ii. The argument from utility. Many theorists have found it tempting to equate the utilitarian argument for democracy with the argument advanced by the relativist.¹¹⁵ But we should resist such a

114. Conventionalists might admit that legislation enacted through a process of representative democracy can, at most, evidence moral facts (i.e., the beliefs of the majority). But they might argue that legislation enacted by direct democracy (by which all citizens vote for or against all proposed enactments) is constitutive of what is moral — for it truly *is* the majority opinion. They thus might conclude that judicial departures from legislative decisions born of referenda or initiatives are necessarily in error.

Such a claim would again be false. The outcome of a vote taken by all members of a community cannot be thought by a conventionalist to do anything more than evidence the fact that the majority believes a particular social arrangement to be moral. It is the belief of a majority — not the manifestation of that belief in a vote — that counts as the moral fact of the matter. Because voting errors can be made and beliefs can change, a judge might rightly find legislation to be in error.

115. See, e.g., JOHN C. GRAY, *THE NATURE AND SOURCES OF THE LAW* 12, 19 (Roland Gray ed., 2d ed. 1921); ROBERT S. SUMMERS, *INSTRUMENTALISM AND AMERICAN LEGAL THEORY* 43-44, 48-49 (1982); Herman Oliphant, *Current Discussions of Legal Methodology*, 7 A.B.A. J. 241 (1921); see also Martin P. Golding, *Realism and Functionalism in the Legal Thought of Felix S. Cohen*, 66 CORNELL L. REV. 1032 (1981) (discussing Cohen's derivation of

temptation. While relativism draws its intuitive strength from metaethical skepticism, utilitarianism draws its strength from metaethical realism. Its claim is that utility should be maximized whether the majority believes that it should or not: hence, majority sentiment is not constitutive of the content of morality.¹¹⁶

James Mill (John Stuart Mill's father) most famously articulated the utilitarian's reason to accord democratic results instrumental value. His argument runs as follows. First, "the concern of Government . . . is to increase to the utmost the pleasures, and diminish to the utmost the pains, which men derive from one another . . ."¹¹⁷ Second, individuals themselves are the best judges of what brings them pleasure and pain. That the majority prefers some course of conduct is thus compelling evidence that this course of conduct will in fact reflect what the utilitarian takes to be moral — "the greatest happiness of the greatest number."¹¹⁸ Finally, because democracy enables the will of the majority to trump the will of the minority, it accomplishes, on utilitarian grounds, maximally good results. Thus, according to Mill, committed utilitarians should favor democracy because it provides the optimal means of tabulating the preferences that are to be maximized under a utilitarian theory of morality. One is more likely to achieve maximal utility by abiding by the will of the majority than by abiding by any other decision procedure.

Because most utilitarians, including James Mill, do not think that individuals are infallible in their judgments about what will bring them pleasure and reduce their pain,¹¹⁹ they must admit the theoretical possibility of instances in which others may assess an individual's prefer-

utilitarianism from relativism); Hurd, *supra* note 23, at 1439-40 (discussing why derivation of utilitarianism from relativism is confused); Michael S. Moore, *The Need for a Theory of Legal Theories: Assessing Pragmatic Instrumentalism*, 69 CORNELL L. REV. 988, 990 (1984) (reviewing SUMMERS, *supra*) (critiquing Summers' conflation of relativism and utilitarianism).

116. To claim that judges ought to maximize the satisfaction of preferences requires one of two claims: Either one must maintain that there is at least one objective (nonrelativized) moral maxim that governs adjudication — the maxim that preferences should be satisfied — or one must combine the claim that all moral maxims are relative to subjective beliefs with the empirical claim that everyone shares the belief that preferences should be satisfied. Because relativists are typically relativists because they find the empirical persistence of profound moral disagreement to be convincing evidence that there are no objective maxims, it is extraordinarily difficult for them to advance the latter claim. Those who derive utilitarianism from relativism thus appear committed to the former self-contradictory claim.

117. James Mill, *Essay on Government*, in *DEMOCRACY*, *supra* note 106, at 43, 44.

118. *Id.* at 43.

119. Though typically not thought to be a utilitarian of the Millian sort, Rousseau captured this common utilitarian assumption when he said: "Our will is always for our own good, but we do not always see what that is; the people is never corrupted, but it is often deceived, and on such occasions only does it seem to will what is bad." JEAN JACQUES ROUSSEAU, *The Social Contract*, in *THE SOCIAL CONTRACT AND DISCOURSES* 1, 26 (G.D.H. Cole trans., 1950). See MILL, *supra* note 5, at 252.

ences more accurately than the individual herself. But this concession opens the door to the possibility that in certain, albeit rare, instances, individuals may judge what is in the interests of a majority more accurately than the majority itself. In such cases, the individual does the right thing by setting aside the will of the majority in favor of the maximal happiness of the majority. Insofar as a judge does precisely this when she *justifiably* disobeys a legislative rule that requires the punishment of a justified offender (who *eo ipso* also rightly recognized that the interests of the majority favored disobedience of a legislatively enacted conduct rule), the utilitarian cannot complain about judges who justifiably acquit justifiably disobedient citizens.

Utilitarianism would generate an exclusionary reason for judges to obey legislative rules that fail to cohere with a majority's true preferences only if it assumed the form of rule utilitarianism, and only if a rule barring all acts of judicial disobedience would produce more utility than would some alternative rule. Because there are well-known reasons to suspect that rule utilitarianism contradicts utilitarianism,¹²⁰ there are reasons to suppose that utilitarians cannot generate rules that provide judges with exclusionary reasons to decide cases in ways that fail to maximize utility. The most that they can generate are rules of thumb. They can justifiably enforce these rules against judges who rightly calculate that such rules should be disobeyed only if they calculate that the punishment of the justified is necessary to deter a disproportionate amount of unjustified disobedience. Absent erroneous judicial disobedience, this rationale for punishing the justified would be unavailable. As such, utilitarianism, if defensible, can at most provide a pragmatic reason to punish justified judicial offenders.

iii. *The argument from institutional competence.* Even if one rejects utilitarianism, and thus rejects Mill's view that democracy yields right results because it accurately reflects what will provide for the greatest happiness of the greatest number, one might nevertheless suppose that judges were ill-equipped to second-guess legislative decisions. That is, one might plausibly think that on any moral theory, not just that of utilitarianism, the institutional constraints imposed on judges prevented them from making accurate moral assessments.

There is a well-rehearsed set of considerations that vindicates the suspicion that judges are institutionally ill-situated to make accurate, all-things-considered moral decisions.¹²¹ Judges are insulated from

120. For his now-famous argument that rule utilitarianism must give way to act utilitarianism in order to be genuinely utilitarian, see LYONS, *supra* note 5.

121. For further discussion of these considerations, see BENJAMIN N. CARDOZO, *THE NATURE OF THE JUDICIAL PROCESS* 113 (1921); RONALD DWORKIN, *TAKING RIGHTS SERIOUSLY*

the political arena and therefore out of touch with the concerns of those whom law is to serve — concerns that on any moral theory are likely to be evidential of what is moral, if not constitutive of it. Their appointment and lengthy tenure in office renders them unaccountable to those whom their decisions affect. Their understanding of the social choices with which their decisions must cohere is limited by the accidental manner in which those choices are presented on their adjudicative agenda. The constraints placed on their factfinding facilities by hectic court schedules and restrictive rules of evidence preclude them from the sort of debate and discussion that freely informs the decisions of legislators. The requirement that they limit their considerations to those issues that arise in the course of adjudicating particular disputes precludes their orderly investigation of all the factors that speak to the pursuit of a social policy. Courts are therefore systemically impoverished in their ability to amass the information relevant to all the things that they ought to consider when called upon to make all-things-considered moral judgments.

Legislators, on the other hand, are not so impoverished. As a group they represent, at least better than does a single judge, the interests of those whom law affects. As such, they are in touch with and can make known the needs and concerns that ought to inform legal decisionmaking. Their short tenure keeps them accountable to those they represent and thus keeps their representation genuine. They possess extensive factfinding resources, virtually unlimited time for debate and discussion, and an open calendar as to the issues that they may consider and the order in which they may consider them. The institutional process by which legislators formulate and enact social policies prevents individuals from accomplishing legislative results that are calculated only to advance personal interests.¹²² As a result of these factors, legislators are, as a group, more likely to make accurate judgments about the social policies that ought to be pursued.

It is clear that these observations provide a reason to punish disobedient judges only if their disobedience fails to reflect the balance of reasons for action. Insofar as these observations should prompt judges to recognize the asymmetry of information that exists between themselves and legislators, they should prompt judges to defer to legislative

22-28 (1977); SARTORIUS, *supra* note 16, at 175-76; Moore, *supra* note 76, at 314-15; Harry H. Wellington, *Common Law Rules and Constitutional Double Standards: Some Notes on Adjudication*, 83 YALE L.J. 221 (1973); see also *Southern Pac. Co. v. Jensen*, 244 U.S. 205, 221 (1916) (Holmes, J., dissenting) (claiming that judicial legislation must be "confined from molar to molecular motions").

122. For further discussion of these legislative "strengths," see Hurd, *Sovereignty in Silence*, *supra* note 43, at 1010-15.

judgments in instances of doubt. If the considerations of institutional competence that make such deference advisable also ensure that judges will probably fail to appreciate the asymmetry of information that exists between themselves and legislators, these considerations provide compelling reasons to think that judicial disobedience of legislative rules will, in most cases, constitute error. But when judicial departures from legislative rules are not erroneous, considerations of institutional competence only provide a reason for surprise; they do not provide a reason for punishment. The most that considerations of institutional competence may provide in such cases is a reason to think that *other* judges are likely to err. If institution designers must punish justifiably disobedient judges in order to inculcate an appropriate degree of deference on the part of judges who would be unjustified in disobeying legislation, then institution designers may have a role-relative reason to punish justified judicial offenders. But such a reason is, again, only pragmatic. In the absence of judicial errors concerning how much deference is due to legislative judgments, considerations of institutional competence fail to provide institution designers with any reason to punish justifiably disobedient judges.

iv. The fear of tyranny. It has been commonly said that to delegate decisionmaking powers to a minority is to invite tyranny. In Lord Acton's famous words, "power corrupts, and absolute power corrupts absolutely." It is for this reason that classical writers maintained that "[a]ll the difficult questions of Government relate to the means of restraining those, in whose hands are lodged the powers necessary for the protection of all, from making bad use of it."¹²³

Alexander Hamilton's famous solution to these difficult questions was twofold: (1) to separate legislative, executive, and judicial powers so that each might check and balance the powers of the others, and (2) to invest the legislative powers in a democratic body that internally checks the ability of individuals to pursue self-interested ends. Because each governmental branch requires for its purposes the powers accorded to the others, each governmental branch is constrained by the others in its ability to achieve its own ends. And because each individual in a democratic legislature requires for her purposes the powers accorded to other individuals, each individual is constrained by others in her ability to act self-interestedly. Under Hamilton's scheme, structural pluralism is the answer to the threat of tyranny.

The executive not only dispenses the honors but holds the sword of the community. The legislature not only commands the purse but prescribes the rules by which the duties and rights of every citizen are to be regu-

123. Mill, *supra* note 117, at 44.

lated. The judiciary, on the contrary, has no influence over either the sword or the purse; no direction either of the strength or of the wealth of the society; and can take no active resolution whatever. It may truly be said to have neither FORCE nor WILL but merely judgment; and must ultimately depend upon the aid of the executive arm even for the efficacy of its judgments.¹²⁴

Yet the arguments so far advanced in this article have failed to suggest that there is a principled basis upon which judges can be constrained from substituting their *will* for that of the legislature in instances in which they rightly conclude that the legislature is in error. Does this not invite tyranny by a minority? Hamilton certainly thought so:

It can be of no weight to say that the courts, on the pretense of a repugnancy, may substitute their own pleasure to the constitutional intentions of the legislature. . . . The courts must declare the sense of the law; and if they should be disposed to exercise WILL instead of JUDGMENT, the consequences would equally be the substitution of their pleasure to that of the legislative body. The observation, if it proved anything, would prove that there ought to be no judges distinct from that body.¹²⁵

Hamilton's willingness to abolish the judicial branch before tolerating judicial disobedience rests squarely on the presupposition that, if judges are licensed to break the law when they deem it in error, they will declare it to be in error when doing so advances their own interests. They will assume a "*pretense* of repugnancy" in instances in which no repugnancy in fact exists. In short, they will abuse their power.

Recognition of this fact provides a reason to be suspicious of judicial disobedience, for it provides a reason to suppose that such disobedience is motivated more by self-interest than by a concern for what is genuinely in the best interests of the community. It provides a reason to think that judicial decisions that depart from legislative decisions also depart from the real balance of reasons for action. But despite Hamilton's hope, the fear of tyranny provides no reason to punish a judge who has *accurately* assessed that the balance of reasons for action favors disobedience. At most, it provides a reason to think that, if such a judge is not punished, others will be tempted to disguise their unjustified disobedience as justified disobedience — to advance disingenuous reasons for disobedience in the hope that they will escape punishment for acts that serve self-interest rather than morality. This reason to punish justified judicial offenders is probably, as a pragmatic

124. THE FEDERALIST NO. 78, at 465 (Alexander Hamilton) (Clinton Rossiter ed., 1961).

125. *Id.* at 468-69.

matter, a good one. But contrary to what many defenders of structural pluralism might have thought, it is not a principled one. Absent temptations toward tyranny, it provides no reason to think that those who assume the constitutional perspective would be justified in punishing justified judicial offenders.

Each of the arguments canvassed so far portrays the value of democracy in instrumentalist terms. Each supposes that democracy is important because it produces decisions that are more likely to be accurate than are the decisions that other decisionmakers reach. Because all of these arguments value accuracy over democracy, none provides a reason to punish a judge who accurately concludes that a democratic legislature is wrong in requiring the punishment of a justified private offender. The most that these arguments provide are reasons to think that any given act of judicial disobedience is more likely than not in error. They may therefore furnish the foundations for an argument from error that licenses the punishment of the justified as a means of curbing unjustified disobedience.

b. Internalist theories of democracy. According to the internalist, democratic decisionmaking is intrinsically valuable. Thus, the fact that a democratic legislature has reached a particular decision functions as an additional reason for judges to comply with that decision. When legislation has been passed on a particular subject, judges have a reason for action that the legislature did not have: the fact that a democratic body has spoken.¹²⁶ This fact must be added to the antecedently existing reasons for obedience (which were applicable to the legislature's decision) and weighed against the antecedently existing reasons for an alternative course of conduct (which were also applicable to the legislature's decision). If the inherent value of democracy provides a weighty reason for obedience, it may tip the balance of reasons for judicial action in favor of an action that would have been grievously erroneous on the antecedently existing balance of reasons for action. Hence, the internal value of democracy may make it right for judges to enforce a law that it was wrong for the legislature to enact.

i. The argument from participation. Carole Pateman has argued that democracy is important because political participation fosters important qualities of personal character.¹²⁷ Following Rousseau and

126. Those who advance an internalist theory of democracy are committed to the thesis that laws enacted by a democratic legislature possess influential authority. See *supra* note 42 and accompanying text. Such laws create new first-order content-independent reasons for acting as the legislature commands that must be added to the balance of antecedent reasons for so acting. See *supra* note 38.

127. See CAROLE PATEMAN, PARTICIPATION AND DEMOCRATIC THEORY 22-44 (1970).

John Stuart Mill, Pateman argues that there is an "interrelationship and connection between individuals, their qualities and psychological characteristics, and types of institutions; . . . that responsible social and political action depends largely on the sort of institutions within which the individual has, politically, to act."¹²⁸ Her argument proceeds as follows. First, as Mill put it, "the general mental advancement of the community, including under that phrase advancement in intellect, in virtue, and in practical activity and efficiency," depends on self-determination.¹²⁹ Second, while a benevolent despot might better achieve decisions that are in the greatest interest of the community, such a governor thwarts self-determination, and thus prevents citizens from developing character traits that are crucial to their intellectual and moral development.¹³⁰ Finally, because democracy alone provides for self-determination, democracy alone guarantees intellectual advancement and the attainment of moral virtue. As Pateman puts this conclusion:

[T]he justification for a democratic system in the participatory theory of democracy rests primarily on the human results that accrue from the participatory process. One might characterise the participatory model as one where maximum input (participation) is required and where output includes not just policies (decisions) but also the development of the social and political capacities of each individual, so that there is "feedback" from output to input.¹³¹

Pateman's participatory theory provides a reason to think that democracy is internally valuable. If valid, it provides judges (and presumably citizens) with a new reason to comply with democratic results — a reason over and above the reasons for compliance that existed prior to the democratic decision to demand such compliance. If a course of conduct is valuable just because it has been democratically willed, then a judge who fails to add the fact of democratic enactment to the reasons to pursue that course of conduct fails to assess accurately the balance of reasons for action. If her ensuing decision to disobey the law would have been different had she added the value of democratic participation to the reasons for obedience, then the judge ought to be punished, for her disobedience is unjustified.

While Pateman's argument provides a reason to think that democracy is valuable, it does not provide a reason to think that the results

128. *Id.* at 29.

129. *Id.* at 28-29 (quoting JOHN S. MILL, *Representative Government*, in *UTILITARIANISM, ON LIBERTY AND CONSIDERATIONS ON REPRESENTATIVE GOVERNMENT* 171, 195 (H.B. Acton ed., 1972) (1861)).

130. *Id.* at 29.

131. *Id.* at 43.

of democracy should be considered exclusionary. Pateman would admit that if democracy failed to advance individual virtue, moral action, and civic-mindedness, it would lack anything other than instrumental value. And one might plausibly argue that a democracy will fail to advance morality unless and until the individuals who participate in it are themselves, to some degree, moral. Democracies are, at least in principle, capable of producing gravely immoral results, and it is unclear how moral virtue might be fostered in a system that requires deeply immoral conduct. While one might argue that immorality is unsustainable¹³² and that the democratic legislation of gravely immoral policies will trigger the democratic change of those policies, it appears equally plausible that immorality breeds immoral character traits that make such change unlikely.¹³³ Hence, it might be the case that for democracy to foster virtue, it must require a certain amount of moral conduct, or it must not, at least, require gravely immoral conduct. Thus, practical reason will compel a judge confronted with gravely immoral legislation to weigh the value of participation against the immorality of the legislative decisions that result from that participation. If the judge concludes that those results will foster greater vice than virtue, the participatory theory of democracy will itself require that she set those results aside.

Once again, then, the value of democracy fails to provide a principled reason to think that judges should be punished for genuinely justified disobedience. As always, the fear that the systemic acquittal of justifiably disobedient judges will foster unjustified judicial disobedience serves as a reason to punish justified judicial offenders. But such a reason is a pragmatic one premised on the fear of disproportionate judicial error.

ii. The argument from autonomy. The argument from autonomy rests on the claim that autonomy is inherently valuable, and, hence, that a decisionmaking process that sums autonomous choices is inherently valuable. Because democracy performs precisely this function of summing autonomous choices, its results are valuable just because they are democratic. This argument is closely aligned with the argument from participation, and, under some constructions, it appears to collapse into that argument. But it differs from that argument when

132. As Fuller supposed, "coherence and goodness have more affinity than coherence and evil." Lon L. Fuller, *Positivism and Fidelity to Law: A Reply to Professor Hart*, 71 HARV. L. REV. 630, 636 (1958).

133. As Mill maintained, "[c]apacity for the nobler feelings is in most natures a very tender plant, easily killed not only by hostile influences, but by mere want of sustenance . . ." MILL, *supra* note 5, at 252. "[W]ill, like all other parts of our constitution, is amenable to habit . . ." *Id.* at 281.

constructed as follows. Participation theorists value political participation only because they value its results. If a benevolent despot could advance individual virtue and civic-mindedness to a greater degree than democracy, participation theorists would no longer find democratic participation valuable. Autonomy theorists, on the other hand, value autonomy for its own sake. That is, autonomy is not to be equated with moral virtue (even if it is intimately connected to it), and hence, even if the exercise of autonomy fails to advance intellectual and moral virtue, it may nevertheless be valuable. Thus, even if democracy produces results that do not contribute to the social development of citizens, its exercise may still serve as a source of value. There are at least two versions of the argument from autonomy — a strong one and a weak one.

(A.) *The strong version.* The strong version of the argument from autonomy runs as follows. First, for an action to have moral worth at all, the actor must autonomously chose it. That is, autonomy is a necessary condition for the moral worth of an action (albeit not a sufficient one). Only if an individual's act is both voluntary (in the sense that it is not coerced by others) and intentional (in the sense that it is the product of deliberation and choice) does that act have moral value. Under this conception, for example, a financial contribution to others will have moral value — and will thus constitute an act of charity — only if it is a product of individual choice. If coerced by others, the act will lack moral worth. It will function like a tax, rather than a gift; it will have good consequences, but it will not be a good act.

Second, compliance with the law has moral worth only if the law itself is a product of the individual's choice. That is, laws constitute pressure from others — such that compliance with them lacks moral value — unless the individual has autonomously endorsed their enactment.

Third, individuals within a community autonomously endorse the laws that govern their conduct only if they have meaningfully participated in the democratic enactment of those laws. There is a vast literature devoted to what counts as meaningful participation, and because I have dwelt on this literature elsewhere,¹³⁴ I do not propose to devote any time to it here. It is enough to recognize that, if one meaningfully participates in a democracy merely by living within a territory that is ruled democratically, then the rules enacted by the government of that territory count as autonomously chosen rules.¹³⁵ If the election of rep-

134. See Hurd, *supra* note 22, at 1657-61.

135. Residency constituted a sufficient condition of participation for Locke. See JOHN

representatives is required for meaningful participation, then laws that are enacted by representative democracy count as autonomously chosen laws. If casting a ballot is required for meaningful participation, then only laws that are enacted by direct democracy (by referendum or initiative) count as autonomously chosen laws. And if casting a ballot *in favor of* the law is required for meaningful participation, then only laws for which one voted in a direct democracy count as autonomously chosen laws.¹³⁶

Fourth, to the extent that judges disobey democratic enactments and implement their own judgments concerning what citizens ought to do, they deprive citizens of the moral worth of their actions. By enforcing laws that differ from those chosen democratically, they coerce citizens to act, thereby defeating the possibility of moral action by citizens.

Finally, institution designers have a moral reason to punish judges who substitute their own judgments concerning what citizens ought to do for those collectively made by citizens themselves. If morality makes autonomous choice a condition of right action, and if only democratically enacted laws are autonomously chosen, then judges cannot substitute their own judgments for democratic judgments without rendering it impossible for citizens to act morally.

If this version of the argument from autonomy were defensible, then the value of democracy would be effectively exclusionary. Judges could never justifiably disobey democratically enacted laws. If judges could not increase moral conduct by substituting their own judgments for those of a democracy, the separation of powers would be inviolate. Judges would be forced to recognize that even though it might have been morally better for the legislature to have chosen a different law — for the balance of reasons for action in fact favors the pursuit of a different social policy — a judicial substitution of that policy could not accomplish morally better results, for its very imposition would deprive it of its worth. Hence, judicial disobedience of legislative rules would never be justified.

There are at least two reasons to think that this version of the argument from autonomy must fail. The first is that it is self-defeating; the second is that it is false. The argument is self-defeating because, if autonomous choice is a condition of moral action, individuals could

LOCKE, *Second Treatise of Government*, in TWO TREATISES OF GOVERNMENT 365-66 (Peter Laslett ed., 2d ed. 1967) (1714).

136. The anarchist Robert Paul Wolff has defended this very stringent condition. See ROBERT P. WOLFF, IN DEFENSE OF ANARCHISM 12-14 (1970); Robert P. Wolff, *In Defense of Anarchism*, in IS LAW DEAD? 110 (Eugene V. Rostow ed., 1971).

not rationally choose to give themselves laws without threatening the moral worth of their own actions. If they act according to their laws when absent those laws they would not so act, their actions lack moral worth. If they act according to their laws when absent those laws they would act in just the same way, then their laws do no work, and the process of their enactment appears irrational. Thus, on the strong view, the democratic enactment of action-guiding rules would be either immoral or irrational.

The strong version of the argument from autonomy appears false because, even though some acts appear to lose their moral worth if the actors who perform them have not autonomously chosen to do so, many acts (and omissions) do not appear to lose their value just because the actor would not choose to perform them without pressure from others. While charity may cease to be charity if it is coerced, truth-telling still constitutes honesty if coerced. Thus, in many instances, one may not deprive others of the moral worth of their actions by pressuring them to act against their will; on the contrary, one may guarantee their moral action when they might otherwise (autonomously) jeopardize it. As such, the strong version of the argument from participation appears too strong, for it fails to take account of the fact that many acts appear to have moral worth even when they are not autonomously chosen.

(B.) *The weak version.* A more plausible version of the argument from autonomy can be constructed with the help of the theory of autonomy developed by Joseph Raz. According to Raz, lives that are autonomously led have greater moral worth than those that are not:

It is the thought that what we are is, in significant respects, what we become through successive choices during our lives, that our lives are a continuous process of self-creation. . . .

. . . .

. . . We regard the fact that a life was autonomous as adding value to it. We think of our own lives and the lives of others as better for having been developed autonomously.¹³⁷

If lives are better for being autonomously chosen, then the acts that comprise those lives must be better for being autonomously chosen. Thus, the individual who chooses to save a drowning child without the promise of financial reward or the threat of punishment does an act that has greater moral worth than the act of the individual who saves the child because of some threat by others. This suggests, as a first step in the weak argument from autonomy, that the autonomy with

137. Joseph Raz, *Liberalism, Skepticism, and Democracy*, 74 IOWA L. REV. 761, 781 (1989). See generally RAZ, *MORALITY*, *supra* note 36, at 369-99.

which a decision is made adds moral value to that decision, although it is not a necessary condition of the moral worth of that decision.

Second, a decision that in fact coheres with the balance of reasons for action, but is not autonomously made, might have less moral worth than a decision that fails to cohere with the balance of reasons for action but is autonomously made. Thus, the choice to contribute to a softball team might have greater moral worth than the choice to contribute to AIDS research if the former is made autonomously while the latter is exacted through threat of sanction.

Third, insofar as democracy tabulates autonomous choices, democratic results represent choices of maximal autonomy. Therefore, democratic choices possess substantial moral worth. Fourth, even when democratic results fail to cohere with the balance of reasons for action, the value that attaches to those results by virtue of their being chosen by a majority may be sufficient, in many instances, to outweigh the value of a choice that in fact coheres with the antecedently existing balance of reasons for action. In those instances, judges ought to defer to the will of the legislature and refrain from substituting what would in fact be a better policy if the legislature had not enacted the law that it did.

Like the argument from participation, this argument provides judges with a reason to obey democratically enacted rules that should be added to the antecedently existing reasons to do what the rules require. But insofar as this argument rests on Raz's understanding of the value of autonomy, it does not purport to be exclusionary. As Raz argues:

[W]e value autonomous choices only if they are choices of what is valuable and worthy of choice. Those who freely choose the immoral, ignoble, or worthless we judge more harshly precisely because their choice was free. . . . This shows that autonomy does not always lead to the well-being of the autonomous person. It can make his life worse if it leads him to embrace immoral or ignoble pursuits. Autonomy contributes to one's well-being only if it leads one to engage in valuable activities and pursuits.¹³⁸

For Raz, the value of autonomy is asymmetrical. Moral acts chosen autonomously have more worth than moral acts performed accidentally or as a result of coercion. But immoral acts chosen autonomously have no moral worth at all and are thus worth less than moral acts performed accidentally or because of coercion.¹³⁹

If Raz is right about the asymmetrical value of autonomy, then

138. Raz, *supra* note 137, at 781-82.

139. To think otherwise is to think that autonomy is a sufficient condition of moral worth, even if it is not a necessary condition.

practical reason will require that judges assess the morality of democratic results. In the event that they are "immoral, ignoble, or worthless," the judge will be compelled to set them aside and substitute a course of conduct that coheres with the balance of reasons for action. Compliance with the judicial substitution will then have some moral worth, albeit not the degree of worth that it would have if autonomously chosen. Thus, if a judge rightly concludes that democratic legislation requiring the punishment of a disobedient citizen is immoral, ignoble, or worthless, the weak version of the argument from autonomy licenses the judge to break the law and acquit the offender.

Yet again, the only justification that a system designer would have to punish such a judge would be that such punishment was necessary to deter others from breaking the law in instances in which the law is not immoral, ignoble, or worthless. If those who assume the constitutional perspective have reason to believe that judges will fail to give due weight to democratic choices that do not suffer from these faults, and if they have reason to believe that only the wholesale punishment of *all* judicial offenders will prevent such errors, then the argument from error will provide them with a role-relative reason to punish the justified. But, as before, this reason will be a pragmatic one, not a principled one.

Both the instrumental and the internal arguments for democracy and the separation of powers fail to provide principled reasons to punish justified judicial offenders. At most, they provide those who assume the constitutional perspective with conditions that trigger the error-based argument from opportunity discussed in section III.A. That argument can now be understood as follows:

(1) Judges are compelled by the demands of practical reason to assess and act on the balance of reasons for action as they see it. Hence, whenever judges conclude that legislation conflicts with the balance of reasons for action — including in that balance the reasons for obedience provided by concerns for error, self-interested bias, the rule-of-law values, and the inherent values of democracy — judges should break the law.

(2) But judges will make errors about how much they will make errors. And judges will be self-interestedly biased about how much they are self-interestedly biased. If judges attempt to decide cases according to what is best, all things considered, then they will frequently ignore or improperly weigh relevant reasons for and against action. Thus, judges would do better if they could escape the demands of practical reason and blindly obey the law, even when the law demands things (such as the punishment of justified offenders) that are, and ap-

pear to them to be, gravely immoral. Judges therefore face a predicament.

(3) Only institution designers can resolve this predicament. By parsing out matters of policy to the legislature, and by restricting judges, on pain of punishment, to the constitutional review and application of those policies, institution designers can reduce moral error. The threat of punishment will provide the judiciary with a subjective reason to defer to the decisions of the legislature — a reason that in most instances will be sufficiently weighty to tip the balance of reasons for action in favor of complying with the legislature's judgments. Insofar as the opportunities to prevent immoral consequences provide actors with reasons for action, and insofar as the power of punishment provides institution designers with a unique ability to prevent immoral consequences, institution designers have a *prima facie*, role-relative reason to punish judges who justifiably fail to punish justifiably disobedient citizens.

Yet, while the fear of judicial incompetence and tyranny may provide compelling pragmatic reasons for punishing judges who are indeed justified in breaking legislatively enacted laws, it remains to be determined whether there are any principled reasons for such punishment. If judges were not susceptible to error or corruption, would there still be reasons to punish them for not punishing justified offenders? Would those who assume the constitutional perspective still have role-relative reasons to enforce the separation of powers?

2. *The Rule-of-Law Values Revisited*

When contemplating the punishment or impeachment of a particular justified judicial offender, those who occupy the constitutional perspective are forced to take on the judicial perspective as well. For to subject a judge's conduct to scrutiny when punishment is at stake is to judge a judge. The rule-of-law values discussed in section III.B of this Part will therefore enter the balance of reasons that determines the justifiability of decisions made by institution designers. Thus, if judges (as well as other officials or citizens) justifiably rely on the punishment of justified judicial offenders, or depend upon their punishment as a means of preserving the coordination necessary to achieve collective goods such as justice, or require their punishment so as to preserve equality among judges, then constitutional actors have at least *prima facie* reasons of a *principled* sort to punish justified judicial offenders.

If reliance and equality are concerns for institution designers, they may provide principled reasons for punishing judges who have justifiably determined that reliance and equality are insufficient reasons to

punish justifiably disobedient citizens. They may, that is, function as principled foundations for constitutional perspectivalism. If these rule-of-law values do vindicate a morality relative to the constitutional role, then they force us to conclude, once more, that morality is inherently paradoxical, for it requires the punishment of those who do the right thing in refusing to punish those who do the right thing.

I do not propose to consider further the extent to which the rule-of-law values provide a principled basis for constitutional perspectivalism. Instead, let me turn to a consideration that appears to provide a principled account of role-relative morality unique to the constitutional perspective.

3. *The Argument from Institutional Efficiency*

Schauer maintains that even if judges made error-free determinations, institution designers would nevertheless have reasons to encourage them to follow legislatively enacted rules when adherence to those rules would render the process of adjudication more efficient.

When judges in courts of law are channelled by relatively precise rules into deciding cases on the basis of a comparatively small number of easily identified factors (Was the defendant driving faster than 55 miles per hour?; Did the plaintiff become ill after consuming a product manufactured by the defendant?), the entire proceeding is streamlined, requiring less time and evidence than would have been necessary under a more rule-free procedure in which a wider range of factors was open for consideration (Was the defendant driving safely? Was the plaintiff's illness caused by the defendant's negligence?). A rule-based system is consequently able to process more cases, operate with less expenditure of human resources, and, insofar as rule-based simplicity fosters greater predictability as well, keep a larger number of events from being formally adjudicated at all.¹⁴⁰

Schauer insists that we should take care "to treat efficiency as a value independent of the value of simplified procedures in diminishing the number of decision-maker errors."¹⁴¹ That is, concern for efficiency need not collapse into concern for judicial error. This conclusion is true, but it takes some work to see why. It would be clearly true if legislative and judicial decisionmaking produced identical results. Then a system designer would have a reason to allocate decisionmaking power to the legislature if that institution could achieve those results at less expense than could the judiciary. Legislative and judicial results would be identical in two sets of circumstances: (1) if legislatively enacted rules perfectly reflected moral rules (so that they

140. SCHAUER, *supra* note 35, at 147.

141. *Id.*

always determined a decision that accorded with a judge's accurate assessment of the balance of reasons for action), and conversely (2) if the total number of instances in which legislatively enacted rules departed from moral rules were identical in number and weight to the total number of instances in which judges failed to assess accurately the balance of reasons for action.

Neither of these sets of circumstances obtains in the cases that concern us. Consider the first possible set of circumstances. If legislatively enacted rules never departed from the rules of morality, then there would never be circumstances in which a citizen would be justified in disobeying legislatively enacted conduct rules, and there would never be instances in which a judge would be justified in disobeying legislatively enacted decision rules that require the punishment of citizens who break legislatively enacted conduct rules. That is, the problem of justifiably punishing the justified would not arise.

Consider now the second possible reason for identical legislative and judicial results. If legislative decisionmaking inevitably produced the same number of errors as judicial decisionmaking, then error would indeed drop out as a reason to prefer one decisionmaking institution to the other. Because institution designers would have reasons to think that the legislature can reach certain decisions (about matters of policy, for example) more cheaply than can the judiciary (given an identical error factor), efficiency would appear to give institution designers a reason to prefer legislative decisionmaking. That is, if judicial disobedience could not accomplish a cumulative increase in morally right results, institution designers would have a principled reason to punish judicial disobedience. Punishment would deter judges from investing resources in attempts to discover whether legislatively enacted rules are in error when those attempts generate as many errors at greater cost as does the blind application of those sometimes erroneous legislative rules.

But while efficiency might decide between decisionmaking systems that produce errors identical in weight and number, this conclusion falls short of providing us with a reason to think that judges should punish justified judicial offenders in circumstances in which judges are assumed to make no mistakes. The question for our purposes is whether efficiency provides a reason to punish judges for breaking legislatively enacted rules when such judicial disobedience in fact reflects the accurate determination that the legislature's rules are over- or underinclusive. Put more bluntly, does efficiency provide a role-relative reason to punish genuinely justified judicial offenders?

For institution designers to be justified on grounds of efficiency in

punishing a judge who justifiably acquitted a justified offender, the judge would have to have miscalculated the proper amount of judicial resources to expend in deciding to acquit such an offender. Take a specific case. If a judge was justified in acquitting a woman after she killed her abusive husband, then *eo ipso* the judge both accurately determined that the law of homicide was overinclusive in her case (or the law of self-defense underinclusive) and accurately calculated that the arguments from error coupled with the rule-of-law values and the values inherent in democracy fell short of providing sufficient reasons to punish her for her disobedience. For institution designers to be justified in punishing such a judge on grounds of efficiency, they would have to be correct in concluding that the resources spent by this judge exceeded those that should have been spent to reach an accurate decision in such a case. That is, they would have to be right in calculating that the judge spent an unjustified amount of time calculating whether the rule in the case was over- or underinclusive and whether the arguments from error, the rule-of-law values, and the values inherent in democracy licensed such an acquittal.

But if institution designers rightly concluded that this judge overcalculated the likelihood of error, would this not be a reason to think that the judge made a mistake? For if the costs involved in deciding that the battered wife should be acquitted exceeded the costs of punishing such an offender, then, other things being equal, is it not the case that the judge should have followed the legislatively enacted decision rule that required her to impose punishment? Efficiency provides a role-relative reason for institution designers to punish disobedient judges only if it provides judges either with no reason, or with a less weighty reason, to punish disobedient citizens.

There are, I think, *prima facie* grounds to suppose that efficiency provides institution designers with more of a reason to punish disobedient judges than it provides judges with a reason to punish disobedient citizens. This is because scarce resources create a prisoner's dilemma for judges. From the judicial perspective, the costs to a citizen of unjustified punishment are likely to trump the costs involved with the sort of protracted litigation required to evaluate whether the applicable legislative rules are either over- or underinclusive in the citizen's case. Thus, in any given case in which a citizen's punishment is at stake, the balance of reasons will probably favor a judge's discovery and application of the background moral considerations that motivated the legislative enactment of relevant rules, rather than a blind application of those rules themselves.

If all judges in all cases second-guessed legislative decisions, how-

ever, the process of justice would grind to virtual halt, and accurate decisions would be achieved at the cost of prolonged and involved litigation. Insofar as it may be more important to process many cases at the cost of a few errors than to process a few cases without error, judges might accomplish better results if they abandoned their attempt to accomplish error-free adjudication (even in cases involving the potential punishment of a justified offender) and blindly applied rules they knew to be over- and underinclusive.

But just as it appears irrational for a citizen to vote, because the incremental value of casting a single ballot is likely to be outweighed by the value of doing work for charity, so it appears irrational for a judge to refrain from second guessing legislative decisions, because the systemic inefficiency that such a recalculation is likely to produce is small in comparison to the costs to a citizen of applying an overinclusive rule (even when those costs are discounted by the probability that the rule to be applied is not in fact overinclusive). Thus, even if judges recognize that the viability of the judicial system rests on the collective willingness of judges to apply legislatively enacted rules without recalculating the likelihood that those rules are in error, the demands of practical reason will nevertheless compel judges to recalculate the likelihood of legislative error in any case in which they suspect that the costs of such error will outweigh the incremental inefficiency that their recalculation will produce.

Just as the punishment of citizens who justifiably disobey the law may rescue citizens from prisoner's dilemmas that threaten collective goods, so the punishment of judges who disobey the law may rescue them from the prisoner's dilemma that threatens the administration of justice. If the universal recalculation of legislation would contribute to the collapse of the judicial system, and if system designers can prevent that collapse by punishing judges who insist on recalculating legislation, system designers have at least a *prima facie* reason to punish judges who second guess the legislature. Insofar as the blind application of legislatively enacted rules strikes the morally appropriate balance between rightly decided and wrongly decided cases, system designers will have a *prima facie* reason to punish judges every time judges second guess legislation, even when judges are justified in so doing because the rule involved is in fact over- or underinclusive. Hence, the concern for efficiency would seem to provide a role-relative reason for system designers to punish justified judicial offenders.

IV. SEIZING THE SECOND HORN: A DEFENSE OF THE CORRESPONDENCE THESIS

In the previous Part, I canvassed the reasons for thinking that institutional roles provide actors with unique reasons for action — reasons that might tip the balance of reasons for judicial action in favor of punishing a justifiably disobedient citizen, or reasons that might tip the balance of reasons for constitutional action in favor of disciplining a justifiably disobedient judge. If such arguments are persuasive, then we must conclude that the correspondence thesis is false: the justifiability of an action does not determine the justifiability of permitting or punishing the action. A citizen might be justified in breaking the law, but a judge might be justified in punishing him for his disobedience. Alternatively, a judge might be justified in breaking the law and acquitting a justifiably disobedient citizen; but a system designer might be justified in disciplining her for her judicial disobedience.

If the correspondence thesis is false, then the appropriate means of resolving the dilemma with which we began is by abandoning the punishment principle. We should admit once and for all that individuals should sometimes be blamed for acting in ways that are morally blameless, and punished for behaving precisely as they should behave. Such a solution has the virtue of preserving the integrity of our systemic values. If judges are morally licensed to punish justifiably disobedient citizens when liberty, equality, and cooperatively achieved public goods are in jeopardy, then we need not fear that recognition of the moral justifiability of certain acts of disobedience will threaten the rule of law. And if constitutional actors are morally licensed to discipline justifiably disobedient judges when democracy and the separation of powers are at stake, then we need not fear that judges will exercise powers institutionally reserved for the legislature in ways that will adversely affect our commitment to structural pluralism.

But while the punishment of the justified will protect institutional values, it will also encourage actors to do what they are not in fact justified in doing. It will prompt citizens to obey unjust laws that ought to be disobeyed, and it will compel judges to enforce unjust laws that ought to be overturned or ignored. While these results may be of institutional virtue, they violate personal virtue. The punishment of the justified creates personal dilemmas: in some instances, individuals must act immorally to be treated as if they acted morally, and they must suffer the institutional blame associated with immorality if they choose to act morally.

Yet this paradoxical "solution" must be accepted only if the correspondence thesis is false, and the correspondence thesis is false only if

the reasons for action explored in the previous Part are in fact role-relative. While the arguments advanced in the previous Part provide *prima facie* reasons to think that concerns about error, the rule of law, and the separation of powers are not concerns that equally affect the balance of reasons for private, judicial, and constitutional action, there are grounds to think that these arguments rest on one or more confusions. In what follows, I shall try to articulate the nature of these confusions. If I am right, dissolving these confusions restores the viability of the correspondence thesis and vindicates the punishment principle. We must thus seek an alternative solution to the dilemma with which we began.

A. *Consequentialism and the Correspondence Thesis*

1. *Rightmaking Consequences for Consequentialism*

Consequentialists who are inclined to embrace role-relative morality, and thereby abandon the correspondence thesis, must defend the claim that the consequences of judicial action are not consequences that enter into the balance of reasons for private action and that the consequences of constitutional action are not consequences that enter into the balance of reasons for judicial action. This claim should trouble most consequentialists because the consequences of judicial action in a case involving a disobedient citizen are consequences that would not occur but for the citizen's disobedience; and the consequences of constitutional action in a case involving a disobedient judge are consequences that would not occur but for the judge's disobedience. That is, a citizen's disobedience is a cause-in-fact (or but-for cause) of any adverse effects on the rule of law or any disproportionate increase of erroneous acts of disobedience on the part of other citizens that occur by virtue of a judge's decision to acquit that citizen. And a judge's decision to acquit such a citizen is a cause-in-fact (or but-for cause) of any adverse effects on the separation of powers or any disproportionate increase of erroneous acts of disobedience on the part of other judges that occur by virtue of a system designer's refusal to discipline that judge. Thus, for consequentialists to defend the truth of role-relative morality, they must have grounds for maintaining that some consequences that would not occur but for a citizen's actions are consequences that do not affect the consequential calculus that determines the rightness or wrongness of that citizen's conduct. Similarly, some consequences that would not occur but for a judge's decision are consequences that do not enter into the balance of consequences that determines the rightness or willingness of that decision. In short, consequentialists who reject the correspondence thesis must preserve the

central consequentialist thesis that an action is right only if it produces more good consequences than bad,¹⁴² while maintaining that some bad consequences should not be included in the calculation that determines the rightness of some actions.

Three possible approaches might appear available to those consequentialists who do not take this project to be impossible on its face. First, consequentialists might attempt to limit the consequences that determine the rightness of actions to those that are proximately caused by actions. Second, consequentialists might try to limit the consequences that determine the rightness of actions to those that an actor can reasonably anticipate or predict. Third, consequentialists might seek to assign weights to rightmaking consequences according to their relative probability at the time of action, thereby making some consequences more likely for some actors than for others, and thus more weighty reasons for action for some actors than for others. As I shall argue, none of these strategies succeeds. The first depends upon a metaphysically indefensible conception of causation; the second and third confuse the conditions of right action with the conditions of culpability.

a. Proximate cause limitations on rightmaking consequences. Consequentialists might attempt to limit the consequences that serve as reasons for and against an action to those that proximately result from that action. They might then argue that the role-relative reasons for judicial action described in the previous Part are not reasons for action for citizens because they do not represent values proximately affected by private conduct. Similarly, the role-relative reasons for constitutional action discussed earlier are not reasons for action for judges because they do not represent values proximately affected by judicial conduct.

To make out these claims, consequentialists would have to argue that judicial action severs the causal chain between a citizen's disobedience and any increase in the number of erroneous acts of disobedience by other citizens or any adverse effects on the rule of law. Similarly, they would have to claim that constitutional action severs the causal chain between a judge's disobedient acquittal of a disobedient citizen and any increase in the number of erroneous acts of disobedience by other judges or any adverse effects on the separation of powers. In short, judicial and constitutional decisions function as the sort of voluntary intervening acts that make the prior acts of others non-proximate to any subsequent systemic consequences.

142. See *supra* note 5 and accompanying text.

This argument draws on the intuition that currently grounds our general refusal to hold individuals legally liable for consequences that they do not proximately cause. Its implicit claim is that our legal rule reflects a deeper moral truth — that consequences that we do not proximately cause are consequences that do not bear on the rightness or wrongness of our conduct.

Notwithstanding the intuitive appeal of such an argument, it provides an untenable foundation for a consequentialist vindication of role-relative morality. Most significantly, its defensibility rests on the ability to make the notion of proximate causation metaphysically plausible, not just morally plausible. That is, it requires one to defend the claim that voluntary, intentional human acts literally sever causal chains.¹⁴³ The arsonist who voluntarily throws a match on a gasoline spill fully intending its subsequent explosion not only does an act that releases those who contributed to the spill from moral responsibility for the fire but renders their acts causally inert.¹⁴⁴ The judge who voluntarily and intentionally breaks the law by acquitting a disobedient citizen not only does an act that releases the citizen from moral responsibility for subsequent errors by others or subsequent adverse effects on rule-of-law values, but renders that citizen's disobedience causally impotent. Only if the argument is advanced as a metaphysical one, not a moral one, can the consequentialist argue that voluntary acts of judges and system designers release prior actors from causal responsibility, and not just moral responsibility, for all subsequent systemic consequences.

143. [A] voluntary act, or a conjunction of events amounting to a coincidence, operates as a limit in the sense that events subsequent to these are not attributed to the antecedent action or event as its consequence even though they would not have happened without it. Often such a limiting action or coincidence is thought of and described as "intervening"; and lawyers speak of them as "superseding" or "extraneous" causes "breaking the chain of causation."

H.L.A. HART & TONY HONORÉ, *CAUSATION IN THE LAW* 71 (2d ed. 1985).

According to Hart and Honoré, a voluntary action capable of breaking the chain of causation is an action that is intentional, uncoerced, and performed with knowledge of its likely consequences. *Id.* at 76-77. A coincidental conjunction of events capable of breaking the chain of causation occurs "whenever the conjunction of two or more events in certain spatial or temporal relations (1) is very unlikely by ordinary standards and (2) is for some reason significant or important, provided (3) that they occur without human contrivance and (4) are independent of each other." *Id.* at 78.

144. As Hart and Honoré describe it:

A throws a lighted cigarette into the bracken which catches fire. Just as the flames are about to flicker out, *B*, who is not acting in concert with *A*, deliberately pours petrol on them. The fire spreads and burns down the forest. *A*'s action, whether or not he intended the forest fire, was not the cause of the fire: *B*'s was.

... Such an intervention displaces the prior action's title to be called the cause and, in the persistent metaphors found in the law, it "reduces" the earlier action and its immediate effects to the level of "mere circumstances" or "part of the history".

Id. at 74.

H.L.A. Hart and Tony Honoré have insisted that such a theory of proximate causation can indeed be defended as a theory of causation, and not just as a theory of moral responsibility.¹⁴⁵ Motivated in part by a desire to make consequentialism defensible, they have sought a means of answering the *reductio ad absurdum* that consequentialist theories have been thought to generate. That *reductio* can be put as follows: If the rightness or wrongness of any act is determined by the balance of its good and bad consequences, and if any act can have a virtually infinite set of consequences, then (1) the morality of virtually all acts is and will remain unsettled (the metaphysical implication), and (2) individuals must calculate an incalculable number of potential consequences in order to reach justified moral decisions (the epistemic implication).¹⁴⁶ For example, insofar as the assassination of Arch Duke Ferdinand of Austria triggered the First World War, and insofar as that War continues to have wide-ranging effects today, the morality of that assassination remains, by this *reductio*, indeterminate, and its practical rationality impossible to assess.

As Hart and Honoré recognize, the *reductio* draws its force from the claim that an act is a cause of all consequences that would not

145. Continuing their discussion of the case of the arsonist, Hart and Honoré argue:

If *A* and *B* both intended to set the forest on fire, and this destruction is accepted as something wrong or wicked, their moral wickedness, judged by the criteria of intention, is the same. Yet the causal judgment differentiates between them. If their moral guilt is judged by the outcome, this judgment though it would differentiate between them cannot be the source of the causal judgment; for it presupposes it. The difference just is that *B* has caused the harm and *A* has not.

Id. at 74 (emphasis added).

It is important to note that, while Hart and Honoré explicitly distinguish their theory of causal responsibility from a theory of moral responsibility, they do not distinguish their thesis as a metaphysical one as distinct from a conceptual one. They, like the ordinary language philosophers of their day, equate metaphysical and conceptual analysis. Yet, insofar as they explicitly declare that their analysis is not limited to the linguistic issue of how English speakers ordinarily employ the term "causation," *id.* at 69-70, they explicitly take conceptual analysis to be more than an analysis of the pragmatics of utterance. For this reason, I follow others in describing their theory of causation as a metaphysical one. See Sanford H. Kadish, *Complicity, Cause and Blame: A Study in the Interpretation of Doctrine*, 73 CAL. L. REV. 323, 334-35 & n.11 (1985). But see Judith J. Thompson, *Causality and Rights: Some Preliminaries*, 63 CHI.-KENT L. REV. 471, 473 (1987).

146. Moral philosophers have insisted . . . that the consequences of human action are "infinite": this they have urged as an objection against the Utilitarian doctrine that the rightness of a morally right action depends on whether its consequences are better than those of any alternative action in the circumstances. "We should have to trace as far as possible the consequences not only for the persons affected directly but also for those indirectly affected and to these no limit can be set." Hence, so the argument runs, we cannot either inductively establish the Utilitarian doctrine that right acts are "optimific" or use it in particular cases to discover what is right.

HART & HONORÉ, *supra* note 143, at 69 (quoting ROSS, *supra* note 22, at 36) (footnote omitted). Thus, as Fumerton has maintained, "[s]ince the consequences of our actions go on and on (far beyond any point we can foresee), one can certainly sympathize with the actual consequence generic utilitarian, Moore, when he despaired of ever knowing what we ought to do." FUMERTON, *supra* note 3, at 105 (citing GEORGE E. MOORE, *PRINCIPIA ETHICA* (1903)).

have occurred but for that act.¹⁴⁷ If this claim is defeasible, so too is the reductio that threatens the coherence of consequentialism. Such a claim is defeasible if the only causes of an event are proximate causes. On Hart and Honoré's theory, acts cease to be proximate and, hence, cease to be causes at all, when they are displaced either by the voluntary, intentional acts of another or by a coincidental conjunction of natural events.¹⁴⁸ On this analysis, one who spills gasoline is in no way a cause of a subsequent fire that is started by an arsonist who drops a match on the spill or by an unprecedented and unforeseeable bolt of lightning that strikes the spill.

Yet it is difficult to claim that such a theory of proximate causation is anything more than a thesis about the justified parameters of moral responsibility. Voluntary, intentional human acts and coincidental conjunctions of natural events would both have to be uncaused causes in order to sever causal chains so as to render causally inert many acts and events that are but-for causes of subsequent consequences. Because we have good reasons to think that there are no uncaused causes and that, even if there were, intentional human acts and statistically infrequent occurrences would not be among them, there appears little promise to the suggestion that a proper theory of (proximate) causation could adequately ground a consequentialist defense of role-relative morality.

This conclusion resurrects the reductio that implies that consequentialists are committed to the thesis that because even the most trivial actions have an infinite set of consequences, the morality of virtually all actions is indeterminate, and the practical rationality of virtually all actions is incalculable. At first blush, consequentialists would seem to have two alternative means of resisting this reductio. They might deny that such a thesis commits them to absurdity, or they might seek an alternative means of limiting the consequences that enter into the balance of reasons that determines right action. As I will argue in the next sections, only the first option is genuinely available. The reductio is not a reductio ad absurdum, because consequentialists need not conclude that the moral indeterminacy of actions precludes the practical rationality of those actions. They need not think a defense of moral indeterminacy commits them to the claim that acts cannot be deemed practically rational absent an ability to

147. HART & HONORÉ, *supra* note 143, at 68-69.

148. See *supra* note 143. As Hart and Honoré argue, these factors function as " 'new actions' (*novus actus*) or 'new causes,' 'superseding,' 'extraneous,' 'intervening forces'" When " 'the chain of causation' is broken" the initiating action is " 'no longer operative,' 'having worn out,' *functus officio*." Hart & Honoré, *supra* note 143, at 74 (citing *Davis v. Swan Motor Co.*, 2 K.B. 291, 318 (1947)).

assess all their consequences. Because consequentialists can plausibly maintain that practical rationality requires only that actors act on the balance of consequences reasonably available to them, they can defend the determinacy of practical rationality while simultaneously embracing the indeterminacy of morality.¹⁴⁹ And once the thesis of moral indeterminacy is divorced from a thesis about practical rationality, it should not be particularly troublesome. Indeed, it is vindicated by our affirmative answers to counterfactuals of the following sort: If those who assassinated Arch Duke Ferdinand could have foreseen that their assassination would trigger World War I, and if they could have foreseen that World War I would result in Hitler's rise to power and his subsequent execution of six million people, should they have considered these consequences as reasons against the assassination? Because it would be the case that, *if* persons possessed crystal balls that allowed them to foresee *all* the consequences of their actions, we would consider their actions morally blameworthy if, over the long run, they did more harm than good, we must conclude that all consequences of actions bear upon the balance of reasons that determines the rightness of those actions.¹⁵⁰ Therefore, our only concern with moral indeterminacy can be an epistemic one — not a metaphysical one. That the morality of actions may be indefinitely indeterminate is not itself a concern; rather, its recognition simply makes clear the degree of practical uncertainty with which we must make moral decisions. Absent a crystal ball, practical rationality requires substantial guesswork. The morality of our actions (though not their epistemic justifiability) will inevitably rest on what will seem to us to be luck.

b. Epistemic limitations on rightmaking consequences. Consequentialists who are unhappy with the suggestion that morality is a matter of luck might seek alternative limitations on rightmaking consequences. First, they might insist that it is unnecessary to construe proximate limitations metaphysically, rather than morally. Individuals should be held responsible only for those consequences that they proximately cause, but those consequences should be thought a function of the limits of moral responsibility, not of the metaphysics of causation. Because actors cannot prevent what they cannot foresee, it is morally unfair to hold them responsible for consequences of their actions that could not be reasonably anticipated. Hence, right action

149. See G.E. MOORE, *PRINCIPIA ETHICA* 150-59 (reprinted 1956) (1903) (arguing that practical rationality consists of considering alternatives thought of while right action may consist of an alternative that is not, and perhaps could not be thought of).

150. As G.E. Moore put it, "[T]he assertion 'I am morally bound to perform this action' is identical with the assertion 'This action will produce the greatest possible amount of good in the Universe'" *Id.* at 147.

should be defined as action that maximizes foreseeable good consequences or minimizes foreseeable bad consequences. Under this definition, the set of an actor's reasons for action is epistemically limited. The rightness of an individual's action is a function of the consequences that he can reasonably foresee at the time of action, and consequences are reasonably foreseeable if the costs in time, talent, diligence, and resources required to appreciate their relative probability are less than the costs of the consequences occurring discounted by their probability.

Because a citizen who contemplates disobedience is poorly situated to judge whether the adjudication of her case will have adverse effects on others' allegiance to the law or on the systemic protection of liberty, equality, and reliance interests, investing significant resources to research the likelihood of such adverse consequences is probably unreasonable. To assess accurately the likelihood of such consequences, citizens would, in effect, be forced to acquire the expertise of a judge. Similarly, because a judge who contemplates the acquittal of a disobedient citizen is ill-equipped to judge whether the evaluation of her disobedience by system designers will have adverse effects on the obedience of other judges or on the separation of powers, investing judicial resources to research the likelihood of those adverse consequences is probably unreasonable. For again, to calculate such consequences, the judge would have to acquire the knowledge and skills of a system designer. Thus, the consequences that ensue from the adjudication of their disobedience should not be thought to affect the rightness of citizens' and judges' disobedience, because citizens and judges are epistemically constrained in their abilities to foresee those consequences.

According to this argument, the consequences that provide a judge with reasons for action do not provide a citizen with reasons for action, because, even though they would not occur but for the citizen's (disobedient) action, they are not consequences that the citizen can reasonably predict. Moreover, the consequences that provide the system designer with reasons for action do not provide the judge with reasons for action because, even though they would not occur but for the judge's (disobedient) action, they are not consequences that the judge can reasonably assess. Morality is role-relative, for different roles provide actors with different predictive powers. Thus, the consequences that enter into the balance of reasons that determines right action within those roles vary substantially. It might therefore be right for one actor to do what it is right for another to prevent or punish.

If successful, this argument would enable consequentialists to defend role-relative morality without embracing a metaphysically suspect theory of causation. But, while this argument escapes the problem that beset the previous attempt to limit rightmaking consequences, it trades one metaphysical confusion for another: instead of confusing the metaphysics of causation, it confuses the metaphysics of right action. By identifying the conditions of right action with those of culpability, it purports to vindicate perspectivalism as a thesis about right action. In fact, it merely establishes the relatively obvious fact that perspectivalism is true as a thesis about culpability, while leaving open the truth of the correspondence thesis as a thesis about right action.

Recall the distinction drawn in Part I between right action and culpable action.¹⁵¹ An actor acts nonculpably if she justifiably believes her actions to be right. Her actions are actually right, however, only if they conform to the objective criteria of our best normative theory. A citizen might well be incapable of predicting that his disobedience will prompt a judge to issue a decision that adversely affects liberty, equality, and cooperatively achieved collective goods. As such, the citizen might act nonculpably when he acts in disregard of the possibility of such consequences. But if his act in fact brings about more bad consequences than good consequences, then his act is wrong, albeit nonculpable. Thus, the consequences that are epistemically available only to the judge may nevertheless be rightmaking for the citizen. Similarly, a judge may well be incapable of anticipating that the evaluation of her disobedience will undermine the separation of powers. As such, she may be nonculpable in failing to consider such consequences when deciding to disobey the law. But if her disobedience indeed leads a system designer to act in ways that threaten structural pluralism, then it may produce more bad consequences than good consequences. Her disobedience may thus be wrong, albeit nonculpable. The systemic consequences that serve as epistemically available reasons for action for the system designer may nevertheless be rightmaking, albeit epistemically unavailable, for the judge.

Those who resist this analysis might invoke one of two rejoinders that purport to provide reasons to collapse the distinction between right action and culpability. The first attempts to demonstrate that absurd results follow from divorcing right action from culpability, for such a divorce allows for the oxymoronic possibility of culpable right action.¹⁵² This argument runs as follows. Suppose that the year is

151. See *supra* text accompanying notes 3-12.

152. For one version of this argument, see FUMERTON, *supra* note 3, at 102.

1930 and the place is Germany. A gunman goes to a local vegetable market where he opens fire, randomly spraying a large crowd of people with bullets. Miraculously, the only shopper he shoots is the young Adolf Hitler — a man altogether unknown to the gunman. Suppose that the gunman acted out of revenge for being fired by a local merchant. He thus acted culpably, for he believed his action to be wrong, or at least unjustifiably believed it to be right. But if culpability is distinguished from right action, and if right action is defined consequentially, his action was right, at least on many versions of consequentialism. By virtue of saving millions of lives, the gunman maximized happiness, preference-satisfaction, virtue, the protection of rights, and so forth, because the millions of people saved from Hitler's forthcoming "final solution" were afforded happiness, virtue, and the exercise of rights that would otherwise have been lost to them. Thus, we must say that the gunman did the right thing.

But is this not an absurd thing to say? Is it not the case that this scenario presents a clear counterexample to the claim that our theory of right action should be divorced from epistemic considerations about the justifiability of an actor's beliefs?¹⁵³ I think not. The claim that the gunman did the right thing only sounds absurd because we let our intuitions about culpability, our presuppositions about the appropriate conditions of punishment, and our equivocal use of the term *right*, swamp our judgment.¹⁵⁴ The gunman was clearly culpable, and on most theories of punishment, culpable action is a sufficient condition of punishment. But to see that the gunman in fact did the right act, it is useful to ask whether he should have done the same thing if he arrived in 1930 as a time traveler, knowing that Hitler was in the crowd, that Hitler would be successful in implementing his so-called "final solution," and that he could kill Hitler without harm to anyone else. Under such circumstances, the consequentialist would have to maintain that the right thing for him to do would be to shoot Hitler.

153. As Fumerton argues:

Because the counterfactual conditionals that define rationality or reasons to act, on [this] view[], ignore the actual epistemic situation of the agent, it would seem to follow that it is possible for conduct to be rational despite the fact that the agent had no reason to believe that its consequences would be better (valued more) than those of its alternatives — indeed, despite the fact that the agent had every reason to believe that the consequences would be far worse. If I were to know all of the information that bears on some decision, I might well act in ways that in fact I have every reason to believe would be disastrous. But surely we want to make the rational/right course of action for me to take a function of my epistemic situation.

Id. at 102-03.

154. Fumerton succumbs to this confusion just because he fails to distinguish between rational (that is, epistemically justified) action and right action. *See id.* He is surely right to make rational action a function of an actor's epistemic situation, but he is wrong to equate rational action with right action and thus to make right action a function of an actor's epistemic situation.

Because this is precisely what the gunman in our original hypothetical did, we must conclude that he did the right thing. But because he did it accidentally, he receives no credit for doing the right thing; indeed, because he in fact thought (or should have thought, given the evidence available to him) that he was doing the wrong thing, he acted culpably and deserves punishment. The concept of culpable right action is no more confused than the concept of nonculpable wrong action. Hence, the possibility of culpable right action gives us no reason to reject the distinction between right action and nonculpable action.

The second move that consequentialists might make in order to defend the claim that the conditions of right action should not be separated from the conditions of nonculpable action draws on the claim that, as a pragmatic matter, the moral evaluation of others turns solely on their culpability. Because the most that we can ask of people is that they act nonculpably, it is pragmatically pointless to work out a theory of right action as distinct from a theory of culpability. Because people lack the hindsight of a time-traveler at the time that they act, they lack complete knowledge of the consequences of their actions. The best that they can do, then, is to make educated guesses about those consequences. What sense does it make to speak of right action when the most that people can aspire to is nonculpable action? Is it not a mere academic exercise (in the pejorative sense) to lay down the conditions of right action when those conditions are at best accidentally achieved? The moral theorist should concentrate on specifying the conditions of culpability, and leave the conditions of right action to God.¹⁵⁵

If the development of a theory of right action is pointless, then the truth of the correspondence thesis as a thesis of right action is irrelevant. We should concentrate on constructing a theory of culpability, and as a thesis of culpability the correspondence thesis is plainly false. Such an argument constitutes a clever strategy for those who seek to reject the correspondence thesis. It vindicates perspectivalism as a thesis about the conditions of culpability, and then maintains that the conditions of culpability exhaust the subject matter of (useful) moral theory. It thus admits the possible truth of the correspondence thesis as a thesis about right action but declares that truth irrelevant.

Tempting as this fallback position might be, however, it too is untenable. The conditions of culpability cannot be specified without possessing an independent theory of right action. Thus the truth of the

155. Elements of this second argument run through Fumerton's defense of an epistemically bounded theory of right action. See, e.g., *id.* at 107-08.

correspondence thesis as a thesis about the conditions of right action is far from pie in the sky. Under the critic's own theory, culpability is a function of belief. An actor acts nonculpably if she reasonably believes her actions to be right. But to know whether her beliefs are reasonable, we must know how close they come to being true. We must thus have a theory of whether her actions are in fact right. Only then can we judge whether the evidence that she employed in formulating her beliefs, and the inferences that she made in arriving at her beliefs, were reasonable. We are able to say that the time-traveler acted nonculpably precisely because we think that Hitler should have been shot (at least if we are employing a consequentialist normative theory), for the fact that Hitler should have been shot makes the time-traveler's belief that he should shoot him reasonable. Likewise, we conclude that the original gunman's actions were unreasonable precisely because he did not know that he would shoot only Hitler. He lacked evidence from which to conclude that he would do the right thing; hence, he did the right thing culpably.

Similarly, only if we possess a theory of right action can we judge whether disobedient citizens and judges have acted nonculpably. That is, only if we know the conditions under which the law should be broken can we judge whether a citizen or judge reasonably believed that it should be broken. If the correspondence thesis is true of the conditions of justified punishment, it is significantly true. It holds that, if a citizen's act of disobedience was right, then the right thing for a judge to do is to acquit the citizen. Hence, the reasonableness of a judge's belief concerning how best to decide the citizen's case must be measured by the degree to which it approaches this truth. Similarly, if a judge is right to punish a citizen for a particular act of disobedience, then that act of disobedience is wrong. Hence, the reasonableness of the citizen's belief concerning whether he should break the law must be measured by the degree to which it approaches this truth. While reasonable belief can depart from true belief, true belief is the goal and measure of reasonable belief. Anyone concerned with the conditions of culpable action must thus admit the necessity of constructing the best possible theory of the conditions of right action. If the correspondence thesis is part of our best theory of the conditions of right action, it bears directly on the project of constructing the conditions of culpability.

c. *Probability limitations on rightmaking consequences.* In light of the above analysis, consequentialists must admit that the reasons for action described in the previous Part as role-relative are in fact reasons for action for all actors; the fact that they are neither proximately af-

fectured by all actors nor epistemically available to all actors is irrelevant to their being rightmaking for all actors. Yet consequentialists might argue that this admission does not defeat perspectivalism. If the weight assigned to these reasons for action can differ as between citizens and judges, then the balance of reasons that determines right action for citizens and judges can differ. While the values that comprise the reasons for action for citizens and judges are identical, their weight may be sufficiently different to justify a judge in punishing a justifiably disobedient citizen. The same, of course, can be said of the possibility that a system designer will be justified in punishing a justifiably disobedient judge.

Why would we think that the weight of systemic concerns might vary as between citizens, judges, and institution designers? The consequentialist's answer must be that these systemic concerns differ in weight because the probability of different actors affecting them differs in value. When the citizen decides to disobey the law, practical rationality dictates that the weight assigned to systemic values be discounted by the probability that the citizen will not be caught or, if caught, that she will not be punished. Only if she is caught, tried, and acquitted will her disobedience result in adverse effects on liberty, equality, and reliance interests; only if her judge is caught, tried, and acquitted will the citizen's disobedience result in adverse effects on the separation of powers. Because the probability that a citizen's disobedience will affect rule-of-law values is likely to be substantially less than 1.0, and the probability that such disobedience will affect the separation of powers is likely to be even less than that, rule-of-law values and constitutional values are appropriately assigned relatively little weight in the balance of reasons for and against private disobedience.

However, at the time the judge acts the citizen has in fact been caught and tried. The probability that an acquittal will affect rule-of-law values is thus substantially closer to 1.0, and the probability that the judge's disobedient acquittal of such a citizen will be detected, evaluated, and approved is significantly increased, so the probability that judicial disobedience will affect constitutional values is greater than was the probability that the citizen's private disobedience would affect those values. Thus, rule-of-law values and constitutional values must be assigned greater weight in the balance of reasons for and against judicial disobedience. By the time that constitutional actors are called upon to review a judge's disobedience, the probability of affecting both rule-of-law values and the separation of powers is quite close to 1.0. Hence, those values assume substantial weight in the balance of reasons for and against punishing the judge. Thus, by virtue of

the fact that the disobedience of different actors will affect systemic values with different degrees of probability, systemic values function as reasons for action that possess different weight for different actors. The differential weight of systemic values may thus make it right for a citizen to disobey the law but wrong for a judge to do so by acquitting the citizen; alternatively, it may make it right for a judge to disobey the law but wrong for system designers to ignore that disobedience.

If defensible, this argument defeats the correspondence thesis. It admits that systemic values are not themselves role-relative, but it provides reasons to think that their weight is. It thus carves out the possibility that the justified might be justifiably punished and so provides a basis for thinking that our original dilemma should be solved by abandoning the punishment principle rather than revising our systemic convictions. Yet this argument succeeds only by smuggling epistemic considerations into the analysis of right action. The argument thus ultimately fails for the same reason that the previous argument failed in its more explicit attempt to limit right action by epistemic constraints. There can be little doubt that, at the time actors act, the consequences of their actions must be assessed according to the relative probability of their occurrence. But this is because actors lack perfect knowledge of precisely what those consequences will be. They are epistemically constrained in their ability to identify the consequences that will actually occur from among the consequences that they think may occur. They are thus practically compelled to consider all possible consequences of conduct epistemically available to them, discounting those consequences by the relative degree to which they have occurred in similar circumstances in the past.

Assuming the truth of physical determinism, however, certain consequences are destined to follow from an action, while others are not. Thus, at the time of action, the actual consequences that will occur possess a probability of 1.0, while the consequences that will not occur possess a probability of 0.0. The citizen who is caught, tried, and acquitted for disobedience at a cost to certain systemic values was thus antecedently destined to affect those values. While she might have been wholly unable to know of the possibility of causing such consequences or to assign a probability of 1.0 to these consequences even if their possibility was known, these consequences nevertheless enter the balance of reasons that determines the rightness of her action (as opposed to the culpability with which the action was performed) without any discount at all. For, on a consequentialist conception of right action, it is the actual consequences of an action that determine the rightness of the action. Because the actual consequences of action in

fact have an antecedent probability of 1.0, they do not differ in their weight as rightmaking reasons for action. The actual probability of affecting systemic values is thus identical for citizens, judges, and system designers, even though these different actors assess that probability differently by virtue of their different epistemic situations.

2. *The Truth of the Correspondence Thesis*

It is now time to examine our position. I have argued that consequentialists cannot defend role-relative morality by limiting the consequences that function as reasons for action to those that are proximately caused by an actor or to those that are epistemically available to an actor. To invoke proximate cause limitations confuses the metaphysics of causation. To employ epistemic limitations confuses the metaphysics of right action. If these conclusions are right, then the consequentialist must admit that *all* consequences of an action enter into the balance of reasons that determines the rightness of that action. Thus, if the systemic consequences of a judge's decision are also consequences of the actions performed by the citizen whose case the judge decides, then those systemic consequences are part of the set of consequences that determines the rightness of the citizen's actions. Similarly, if the institutional consequences of a system designer's decision are also consequences of the adjudicatory result reached by the judge whose case is evaluated by the system designer, then those institutional consequences are among the consequences that determine the rightness of the judge's decision and the rightness of the citizen's decision that caused the judge's decision.

It is crucial to see what follows from this analysis. While a citizen's disobedience might not be as visible as a judge's and hence might not itself generate a disproportionate amount of unjustified disobedience on the part of others, the fact that a judge's subsequent acquittal of such a citizen is a consequence of the citizen's disobedience makes it the case that the disproportionate amount of unjustified disobedience that the judge's highly visible disobedience generates is in fact a consequence of the citizen's disobedience. As such, that consequence not only affects the rightness of the judge's acquittal; it also enters into the balance of reasons that determines the rightness of the citizen's original disobedience. Similarly, while a citizen's refusal to cooperate with a legally established cooperative strategy to a prisoner's dilemma will not itself lead to the dissolution of that strategy (and the subsequent loss of an important collective good), the fact that a judge's subsequent acquittal of that citizen is a consequence of that citizen's disobedience makes it the case that any adverse effects on the cooperative strategy

generated by the judge's disobedience are consequences of the citizen's disobedience. As such, those adverse effects enter into the balance of reasons that determines the rightness of the citizen's disobedience to begin with; they do not just enter into the balance of reasons that determines the rightness of the judge's disobedience. The same sort of analysis establishes that the institutional effects of constitutional decisions are consequences that affect not only the rightness of constitutional decisions but also the rightness of judicial decisions and the rightness of private decisions. While the disobedience of a citizen, and the acquittal of that citizen by a judge, are acts that are by themselves unlikely to affect the efficient separation of governmental powers, the fact that a system designer's subsequent decision to ignore such disobedience is a consequence of both the citizen's disobedience and the judge's disobedience makes it the case that any adverse effects of such a decision on structural pluralism are consequences of both the citizen's disobedience and the judge's disobedience. Those institutional consequences therefore affect the rightness of both the citizen's conduct and the judge's conduct, as well as the system designer's conduct.

Thus, none of the reasons characterized in the previous Part as role-relative are in fact role-relative on a consequentialist understanding of morality. The rule-of-law values of liberty, equality, and reliance are reasons for action for citizens as well as for judges, and the institutional values of structural pluralism are reasons for action for citizens and judges as well as for system designers. While some actors may be epistemically ill-situated to evaluate the degree to which their actions will affect these values, the consequential effects of their actions on these values nevertheless enter into the determination of the rightness of their actions.

We are now in a position to advance on behalf of the consequentialist a tentative and partial answer to the original puzzle. Actors are justified in disobeying the law only if the consequences of their actions, including *all* those systemic consequences that they cannot predict or accurately weigh, favor disobedience. If the consequences of their disobedience, including all those systemic consequences that are epistemically unavailable to them, are worse than the consequences of their obedience, then they are wrongdoers who are not in fact justified in disobeying the law.

This more sophisticated understanding of what it means to be justifiably disobedient according to a consequentialist account of right action lends enormous plausibility to the correspondence thesis. To appreciate this fact, consider again the case of the battered wife who violates the law by preemptively killing her abusive husband after he

threatens her life and the lives of her children. Suppose that the consequences of her disobedience, including the adverse systemic consequences caused by her acquittal and the acquittal of her judge (including all adverse effects on the obedience of others, the rule-of-law values, and the separation of powers), are not as bad as the consequences of her obedience. Under such circumstances, the wife's disobedient action is the right action; and it is right not just for her, but for her judge and her judge's judge. The balance of reasons that determines the rightness of her disobedience is identical to the balance of reasons that determines the rightness of their disobedience. Hence, if her disobedience is right, then their disobedience must be right. Her judge does the right thing in acquitting her, and her judge's judge does the right thing in acquitting him, because included in the calculus that makes her act right is the fact that an acquittal of her by her judge, and an acquittal of her judge by a system designer, will not disproportionately encourage unjustified acts of disobedience on the part of other citizens or judges and will not unduly affect the rule of law or the separation of powers. The rightness of her act thus determines both the rightness of a judge's decision to acquit her and the rightness of a system designer's decision to acquit her judge.

The correspondence thesis thus appears to be vindicated: the rightness of a citizen's disobedience determines the rightness of not punishing that citizen, and the rightness of a judge's disobedience determines the rightness of not disciplining that judge. On a consequentialist account of right action, it appears that the justified cannot be justifiably punished.

Yet, notwithstanding the truth of the correspondence thesis in the sort of case that I have described, it is hasty to conclude that a sophisticated understanding of consequentialism renders the correspondence thesis necessarily true. It appears that there is a special set of limiting cases that confounds our ability to generalize the truth of the correspondence thesis to all cases involving codependent actions.

3. *Limiting Cases: Self-Referential Rightmaking Consequences*

While the above scenario illustrates the plausibility of the correspondence thesis for consequentialism, there appear to be some special cases that limit its generalizability. Unlike their predecessors, these cases do not appear to be generated by conceptual confusions about the metaphysics of causation or the metaphysics of right action. They thus provide a compelling basis for thinking that consequentialism can

incorporate perspectivalism.¹⁵⁶

Suppose the acquittal of a battered wife for the murder of her abusing husband will substantially increase the number of instances in which other women will invoke this "solution" in circumstances in which it is quite unjustified (e.g., when murder is simpler or more lucrative than divorce). Suppose further that such an acquittal will have significant adverse effects on liberty, equality, and certain cooperatively achieved collective goods. Together, these bad consequences outweigh the good consequences that come of her action. While the killing will save three innocent lives, others who copycat the killing in unjustified circumstances will take more than three innocent lives. While the killing will relieve the psychological torment of the battered wife and her children, it will trigger a disproportionate amount of anguish on the part of those whose loved ones are killed unjustifiably, and it will engender a disproportionate amount of insecurity among those who fear that they or those they love will be the targets of future unjustified killings. According to the analysis in the previous section, the battered wife's disobedient action is unjustified: the consequences that function as reasons for and against action favor the decision not to kill her husband.

But suppose that her punishment would alleviate most of the adverse systemic effects of her killing — so many, indeed, that her act would then produce more good consequences than bad consequences. Her punishment would deter other women from killing their husbands in circumstances in which such killings would be unjustified, and it would protect rule-of-law values that would otherwise be adversely affected. Under such circumstances, it appears that her punishment would make her action right. It would make it the case that she saved three innocent lives without causally contributing to the loss of a greater number of innocent lives.

Such a case is deeply puzzling. In the absence of punishment, the battered wife does the wrong thing to save her own life and the lives of her children, for by so doing she causes others to take more than three innocent lives. Thus, an acquittal makes her a wrongdoer. It treats as blameless a person who is morally blameworthy — a person whose punishment would not constitute punishment of the justified. Yet if punished, she is a rightdoer, for her punishment has the happy consequence of making right an act that is otherwise wrong. But, by virtue of making her conduct right, her punishment treats as blameworthy a

156. I owe my recognition of the significance of these limiting cases to lengthy discussions with Larry Alexander. For his excellent analysis of these sorts of cases, see Alexander, *Pursuing the Good*, *supra* note 20.

person who is now morally blameless. For the price of punishment, this actor saves three lives without causing the sort of bad consequences that would make her a wrongdoer deserving of punishment!

By demonstrating the conceptual possibility of actions that are morally justified if and only if punished, this scenario poses a powerful counterexample to the truth of the correspondence thesis as a consequentialist thesis about the conditions of punishment. Given the self-referential nature of the consequences involved, the rightness of a citizen's action does not determine the rightness of not punishing that action. Indeed, the rightness of the citizen's action depends upon its punishment.

Because this case does not smuggle in considerations of culpability or considerations of proximate causation, it cannot be dismissed with previous cases as conceptually confused. But there is a set of reasons to think that such a case is quite special — so special in fact that its possibility need not defeat the general viability of the correspondence thesis for consequentialism.

First, while the conceptual possibility of self-referential consequences defeats the claim that the correspondence thesis is necessarily true as a thesis about the conditions of punishment, it does not defeat the claim that the correspondence thesis is necessarily true as a thesis about the conditions of preventative or permissive actions. Cases involving preventative or permissive actions do not permit self-referential consequences. Hence, the correspondence thesis as it was originally stated remains necessarily true for consequentialism.

Recall that, as originally stated, the correspondence thesis held that the rightness of an action determines the rightness of permitting that action. Cases involving self-referential consequences do not affect the truth of this original thesis. As the previous subsections have made clear, if it is right for an individual to do an act, then, on a consequentialist theory of right action, that act must produce more good consequences than bad consequences, *all consequences considered*. An act that prevents a right act prevents not only its bad consequences but also its good consequences. Thus, if the good consequences of the act outweigh its bad consequences, a preventative act fails to maximize good consequences. Hence, if an act is right, its prevention must be wrong.

Those who doubt that an act and its prevention cannot be self-referential might toy with the following sort of counterexample. Suppose that a preventative act would turn a completed act into a mere attempt. Is it not possible that the attempt would be right but that its completion would be wrong and, hence, that the prevention of its com-

pletion would be right? Consider the case in which a nation builds a nuclear bomb as a means of deterring another nation from waging war against it. Suppose that those who control the bomb would not be justified in deploying it (because, for example, it would annihilate far more civilians in enemy territory than would be lost domestically should war be declared by the enemy nation). But suppose further that, in the event of escalating violence, the only way to convince the enemy nation that the bomb will be used is by commencing the procedures that will deploy the bomb. Is it not both right for one individual to act so as to deploy the bomb and right for another individual to prevent the successful completion of that task?

While this scenario appears to make self-referential the rightness of an act and the rightness of its prevention, it fails to defeat the truth of the correspondence thesis. If an attempt is right, then anyone who thwarts the attempt at the stage at which it is but an attempt acts wrongly. If the completion of that attempt is wrong, then the attempter acts wrongly when he does not abort the attempt so as to prevent its completion. Hence, another's prevention of his attempt at that point will be right. In both cases, the rightness of an act determines the wrongness of its prevention, and the wrongness of an act determines the rightness of its prevention. The correspondence thesis thus looks airtight in any case that involves an act and its prevention or permission.

But punishment is special. While an act of punishment prevents future acts of just the sort for which the actor is punished, it does not prevent that actor's past act. Thus, unlike an act of prevention, an act of punishment has the capacity both to preserve the good consequences of the punished act and to prevent at least some of its bad consequences. Thus, punishment of a battered wife who has preemptively killed her husband does not take away the good consequences of that act (i.e., the lives saved). It also does not take away a number of its bad consequences (i.e., the loss of the husband's life, the grief of the children, the emotional trauma of the wife). But it does take away, or at least reduce, the bad consequences that the act has on future actors and on systemic values (i.e., its precedential effect on others who would kill in unjustified circumstances, and its adverse effects on liberty, equality, and cooperatively achieved collective goods). If punishment reduces the bad consequences of such an act to the point at which the good consequences of the act outweigh its remaining bad consequences, then punishment can make that act right.

Does this conclusion force us to confine the correspondence thesis to acts of prevention and permission? As a conceptual matter, yes.

But this admission need not defeat the consequentialist's defense of the thesis as contingently true of punishment. There are empirical reasons to think both that the correspondence thesis is contingently true in the vast majority of cases involving punishment and that it could and should be made true in all remaining cases.

First, as an empirical matter, the class of cases in which punishment can make an otherwise wrong act right appears very small. It includes only those cases in which most of the bad consequences of a disobedient act result from the institutional adjudication of that act, i.e., cases in which citizens would be justified in breaking the law were it not for the fact that their disobedience will cause judicial publication of their actions. For only in such cases do judges have the ability to affect the bad consequences of citizens' private acts. That is, only when a judge will be causally responsible for the disproportionate number of bad consequences of a citizen's disobedience will she have the power to alleviate those bad consequences so that the citizen's actions produce a disproportionate number of good consequences.

Second, there are reasons to think that this small class of cases would dwindle to zero over the long run if judges refused to punish persons in these cases. If judges consistently refused to punish battered wives who killed their abusing husbands to save their own lives, sufficient decisions would be amassed to make clear both the conditions under which spousal killings are justified and the conditions under which spousal killings are unjustified. With the multiplication of such cases, the number of erroneous killings would dwindle to the point at which otherwise justified killings would not trigger a disproportionate number of unjustified killings. From that point forward, more innocent lives would be saved than lost by acquitting battered wives who killed their husbands as a necessary means of saving their own lives. Thus, from that point forward, punishment would deter more justified killings than unjustified ones. So long as more innocent lives are saved after this point than are lost prior to this point, the right thing for judges to do is to acquit battered wives prior to this point. Put simplistically, while the first acquittal of a battered wife who saves three lives by killing an abusive husband may trigger ten unjustified killings, the second is likely to trigger fewer unjustified killings, for it makes more clear the conditions under which killings are justified and unjustified. And the third is likely to trigger still fewer unjustified killings, as are the fourth and the fifth. At the point at which an acquittal triggers fewer unjustified killings than justified killings, battered wives no longer act wrongly (in the absence of punishment) in killing their abusive husbands to save their lives. From that

point forward, acquittals will encourage more right action than wrong action while relieving right actors of the harm of punishment. Hence, so long as there will be more cases after that point than prior to that point, the right thing for judges to do is to acquit disobedient actors prior to that point.

Thus, just as all consequences enter into the balance of consequences that determines the rightness of a citizen's disobedience, so all consequences enter into the balance of consequences that determines the rightness of a judge's act of punishment. If an acquittal will in the long run maximize good consequences, then the right thing for a judge to do is to issue an acquittal. And in cases in which a citizen's action is deemed wrong because of its bad systemic consequences, her acquittal is probably right: Her acquittal, coupled with the successive acquittals of those like her that will follow out of deference to the value of equality, will reduce over the long run the bad systemic consequences of actions like hers. At some point, those actions will become right, because their good consequences will outweigh their bad systemic consequences. With the multiplication of such cases, the earlier acts of disobedient citizens are made right. And the rightness of those earlier acts makes it right that the citizens who performed them were acquitted.

This argument does not make the correspondence thesis *necessarily* true as a consequentialist thesis about the conditions of justified punishment. It merely suggests that it is *contingently* true as a consequentialist thesis about the conditions of justified punishment. It remains possible to construct cases in which the consequences of certain private acts and the consequences of their punishment remain permanently self-referential. In such cases, the correspondence thesis will remain false. But such cases are likely to be the product of academic fancy. So long as we think that the law can guide persons to act rightly, we have reason to think that persons should be acquitted when their actions are made wrong only by the fact that the law lacks the clarity to succeed perfectly at this task.

It thus appears legitimate to take the correspondence thesis to be part of our best consequentialist theory of the conditions of right action. It follows that a consequentialist morality is not, in principle, paradoxical, for it neither requires nor permits the punishment of the morally justified. If this is the case, then the dilemma with which we began our discussion cannot be resolved by abandoning the punishment principle.

Before we make reference to how consequentialists might best solve that dilemma, it is important to determine whether the corre-

spondence thesis is also part of our best deontological theory of right action. If it is, then deontologists must join consequentialists in their search for a means of squaring our systemic values with a commitment to the punishment principle. If, on the other hand, the correspondence thesis is not part of our best deontological theory, then deontologists will not be committed to preserving the punishment principle. The jurisprudential question of whether to punish the justified in order to preserve the rule of law and the separation of powers will then have to await the resolution of the normative dispute between consequentialists and deontologists.

B. *Deontology and the Correspondence Thesis*

Recall that deontologists deny the consequentialist claim that the rightness of an act consists in its maximization of good consequences.¹⁵⁷ They locate the goodness of an act not in its consequences but in the act itself. According to deontologists, some act-types are intrinsically right and other act-types are intrinsically wrong. Moral action consists in complying with agent-relative maxims that categorically prohibit or require the performance of certain acts. Individuals do not act rightly in violating the conditions of right action so as to maximize the instances in which they or others act rightly. If it is wrong to kill the innocent, then an agent is prohibited from killing an innocent person even if, by so doing, he prevents another agent from killing many innocent persons in violation of the agent-relative prohibition directed at her. If it is right to tell the truth, then an actor does wrong to lie even if the lie will bring about substantially more acts of truth-telling by others. And if it is wrong to punish persons who nonculpably do the right thing, then it is wrong to punish the justified even if doing so will dramatically reduce the instances in which the justified are punished.

In Part I, I suggested that any plausible deontological theory would comply with the correspondence thesis, at least when construed in its original form as a thesis about the conditions of preventative and permissive actions. But this is not because the thesis in its original form is necessarily true of deontology. While denial of the correspondence thesis as a thesis solely about the conditions of preventative and permissive acts constitutes self-contradiction by a consequentialist, it need not constitute self-contradiction by a deontologist. It may not be logically possible for a deontological system to contain contradictory maxims (maxims that simultaneously obligate an agent to do act *A*

157. See *supra* text accompanying notes 10-12.

and not to do act *A*). But it is logically possible for a deontological system to contain simultaneously binding maxims that, as a practical matter, cannot be mutually fulfilled. Hence, in a case of intrapersonal competing maxims, a deontological system might simultaneously obligate a single actor to do act *A* and act *B*, where *A* and *B* cannot both be done in the actor's circumstances. Thus, to recall Sartre's famous dilemma, a deontological morality might simultaneously provide a son with an agent-relative obligation to care for his mother and an agent-relative obligation to join the Free French, when he cannot practically fulfill both obligations. And, in a case of interpersonal competing maxims, a deontological morality might provide one actor with an agent-relative obligation (or permission) to do an act while giving another actor an agent-relative obligation (or permission) to prevent that act. Thus, it is logically possible for a deontological system to provide an actor with an agent-relative permission to kill another if necessary to defend her life while providing others with agent-relative obligations to prevent killings, including those that are done in self-defense.

It is even logically possible for a deontological system to contain self-referential maxims, thus generating the sort of puzzle that I discussed in the previous section. A citizen might be accorded an agent-relative permission to act only if she will be punished for so doing. A judge might simultaneously be subject to agent-relative obligations to punish the unjustified and acquit the justified. Under such a deontological system, the judge's punishment would make the citizen's action right, and so violate the prohibition against punishing the justified. But the judge's refusal to punish the citizen would make the citizen's action wrong and so violate the obligation to punish the unjustified. Under such a deontology, the rightness of one actor's conduct would thus make the other actor's conduct wrong.

I argued in Part I that, while the agent-relativity of competing interpersonal norms prevents them from being contradictory, it does not prevent their mutual assertion from being highly untenable. Competing interpersonal maxims would make us moral gladiators. Our successful fulfillment of our moral obligations would thwart the successful fulfillment of others' moral obligations, and their successful fulfillment of their moral obligations would thwart the successful fulfillment of our moral obligations. We would be persistently prevented from doing what we are permitted to do, and we would persistently prevent others from doing what they are permitted to do.

If it is the case that a plausible deontological theory would not pit actors against one another in moral combat by permitting or obligating some actors to prevent others from doing what they are permitted

or obligated to do, then it is seemingly the case that such a theory would not permit or obligate some actors to punish others for doing what they are permitted or obligated to do. That is, if a plausible deontological theory reflects the correspondence thesis in its maxims concerning preventative and permissive acts, it is reasonable to suppose that it also reflects the correspondence thesis in its maxims concerning acts of punishment. For a system that requires the punishment of the justified makes the conditions of right action gladiatorial. Citizens who have done the right thing must in some cases be treated by judges as if they have done the wrong thing. For judges to accord citizens moral success in such cases constitutes moral failure on the part of those judges. A judge's moral success thus requires her to treat morally successful citizens (citizens who have been justifiably disobedient) as if they were guilty of moral failures.

In what follows, I propose to explore the extent to which a plausible deontological theory is in fact committed to a nongladiatorial conception of the conditions of right action. Specifically, I propose to take up four sorts of cases that suggest the simultaneous rightness of competing interpersonal maxims. If deontologists would resist the suggestion in these cases that persons can be obligated to compete for moral success, then we have some cause to think that the correspondence thesis is naturally compatible with the most plausible deontological theory. We thus have reasonable grounds to maintain that among the agent-relative obligations that bind actors under a deontological theory is the obligation not to punish the justified. If, on the other hand, deontologists would embrace gladiatorial maxims in these cases, then we have some reason to think that they may be untroubled by the prospect of punishing the justified, at least in cases analogous to these test cases.

1. *Sporting Competitions*

The first cases that might give deontologists pause in their incorporation of the correspondence thesis are cases of competitive sport.¹⁵⁸ Consider, for example, the sport of boxing. Is it not the case that in a boxing ring it is right for each boxer to prevent what it is right for the other to do? That is, assuming for the moment that the sport of boxing is moral, is it not both morally acceptable for one boxer to throw a punch and morally acceptable for the other boxer to block that punch? It would seem that this is rather the point. But, as a result, it would

158. I owe the incorporation of this discussion to Samuel Freeman, who fruitfully suggested during a colloquium discussion that boxing matches pose *prima facie* counterexamples to the correspondence thesis.

seem that the maxims that apply in the boxing ring defy the correspondence thesis, for the rightness of one competitor's action does not determine the rightness of the other's permission of that action. And the same can be said of the maxims that guide players in other sports: in football it is both right for a nose guard to block a punt and right for a center to block the nose guard from so doing; in basketball it is both right for a guard to shoot and right for an opposing player to block the shot.

Indeed, it would seem that what makes many sports sporting is that their rules suspend the correspondence thesis. But this fact, far from defeating the truth of the correspondence thesis as a metarule of deontological morality, would seem to constitute the exception that proves that rule. What is special about sports of the sort I have mentioned is that a player's *consent* to participate in them waives the application of maxims that would otherwise apply both to his conduct and to that of others. A football player's consent turns what would be a brutal battery into an admirable tackle. It thus suspends maxims that would otherwise require the sort of peaceful interaction that makes for poor spectator sport — maxims that appear to embody the correspondence thesis. Absent consent, deontological norms would surely prohibit the sort of conduct that passes for sport. That is, they would surely prohibit others from punching an unconsenting actor, or running headlong into her, or smashing balls at her face, or jumping up and down in front of her to impede her progress. Thus, while competitive sports are competitive precisely because they impose on participants competitive maxims of action (and so violate the correspondence thesis), they do so justifiably only because they do so by the consent of the participants. It is the consent of the participants that makes it morally acceptable for them to do to each other what they are not justified in doing to others.

Sports thus demonstrate that acts of consent might plausibly suspend the correspondence thesis within a deontological morality. This insight has important implications in the context of punishing the disobedient. It suggests that punishment of the justifiably disobedient might be justified if, but perhaps only if, the justifiably disobedient consent to their own punishment.

At least three things must be said about this suggestion. First, as an empirical matter, it is rarely the case that individuals can be said to consent to their own punishment — particularly when they are being punished for acts that they are permitted or obligated to perform.¹⁵⁹

159. *But see supra* note 85 (discussing civilly disobedient actors).

However, this fact might prompt theorists to claim that justifiably disobedient citizens *ought* to consent to their punishment and, hence, that their failure to do so constitutes practical irrationality. Insofar as fully rational, fully informed actors would consent to their own punishment, the punishment of the justified is justified by the hypothetical consent of the justified.

Hypothetical consent is notoriously problematic. We would think it outrageous for a court to take seriously the claim of a rapist that while his victim did not in fact consent to intercourse, she would have consented to it had she been fully rational; hence, his forced penetration did not constitute a rape. Such a case makes clear that hypothetical consent lacks just the element that makes an act of consent morally significant; namely, consent. If hypothetical consent is a morally meaningful construct, it is not because it embodies or reflects consent. Rather, it is because it refers to something that is moral and thus it captures something to which it would be moral to consent.

This raises the third and most significant point. Consent should be thought to alter the normative environment — to grant permissions, confer rights, impose duties, and so forth — only if it is morally justifiable for it to do so. To appeal to an individual's consent in the course of determining whether her consent should be thought to have such power is circular. As Raz has nicely argued:

Consent . . . is an act purporting to change the normative situation. Not every act of consent succeeds in doing so, and those that succeed do so because they fall under reasons, not themselves created by consent, that show why acts of consent should, within certain limits, be a way of creating rights and duties. We cannot create reasons just by intending to do so and expressing that intention in action. Reasons precede the will. Though the latter can, within limits, create reasons, it can do so only when there is a non-will-based reason why it should.¹⁶⁰

Thus, we must establish, independent of arguments from consent, when, and under what circumstances, acts of consent should be thought to legitimate the practices consented to. Only then shall we be in a position to evaluate the suggestion that a justifiably disobedient actor's consent might justify his punishment.

I do not propose to engage in this quite worthy project of determining why, and under what conditions, consent should be thought to grant permissions, accord rights, or engender duties. As others have made clear, there are both instrumental and noninstrumental reasons to accord individuals the normative power to alter rights and duties. Such a power instrumentally expands liberty by creating institutions

160. RAZ, *MORALITY*, *supra* note 36, at 84.

that advance useful goals and enable individuals to make long-term plans. Such a power is also noninstrumentally constitutive of certain sorts of relationships: as Julie Andrews sang it, "Love isn't love till you give it away."¹⁶¹ Common to both an instrumental and a noninstrumental account of the proper significance of consent, however, is the claim that consent makes possible institutions and relations that are themselves morally valuable. But consent does not make institutions and relations morally valuable; it only makes morally valuable institutions and relations possible.

Thus, as a condition of establishing that consent should operate to legitimate a particular practice, one must establish that that particular practice is itself moral or just. We think that we can do that in the case of sports (and it is just because many doubt that we succeed in the case of boxing that that sport is so morally worrisome). We do not think that we can do that in the case of slavery or murder. We therefore do not take consent, however knowingly or voluntarily given, to legitimate a person's enslavement or murder. Can we establish that the punishment of the justified is moral or just? That is precisely the question with which we are concerned.

Just as we discovered that reliance on the punishment of the justified could not be justified unless the punishment of the justified is independently justifiable, so too, consent to the punishment of the justified cannot be justified unless the punishment of the justified is independently justifiable. We thus cannot use the possibility that justifiably disobedient persons would consent to their own punishment to ground a claim that their punishment is consistent with a plausible deontology. We must continue to seek other reasons to think that deontology would permit the punishment of the justified. If we find other reasons to suspend the correspondence thesis in cases of punishment, then the consent of a justifiably disobedient actor may legitimate a particular instance of punishment. It may even make sense to talk about hypothetical consent on the part of justifiably disobedient actors who do not give actual consent to their own punishment. But these conclusions will follow only if a plausible deontological theory justifies the punishment of the justified. It will only do so if the following cases vindicate the suggestion that the maxims for action applicable to citizens, judges, and institution designers can be gladiatorial in content.¹⁶²

161. RICHARD RODGERS & OSCAR HAMMERSTEIN II, *Reprise of You Are Sixteen*, in *THE SOUND OF MUSIC* (1965).

162. The argument advanced here against the claim that consent might suspend the correspondence thesis in cases of punishment also applies to the suggestion that promises can suspend the correspondence thesis in cases of punishment. It thus precludes one from arguing both that the judicial oath of office contains a promise to apply sanctions to justifiably disobedient citizens

2. *Gladiators and Innocent Shields*

A second set of cases that appears to defy the correspondence thesis involves instances in which innocent persons are forced to confront one another in life-threatening combat. Consider the true gladiators — innocent men coerced to duel until death by threats of immediate execution. While the purpose of such confrontation was sporting, the sport lacked the consent of the participants. Nevertheless, was it not right for each gladiator to attempt to take the life of the other while preventing the other from taking his own life? Now consider the more contemporary case of the innocent person who is forced to shield a culpable criminal in a shoot-out with an innocent police officer. In such a case, is it not right for the police officer to defend herself by shooting through the innocent shield in order to kill the criminal? And is it not simultaneously right for the innocent shield to defend herself by shooting the officer (assuming that she can do this, but cannot shoot the criminal who holds her hostage)? In cases of dueling gladiators and innocent human shields, the correspondence thesis appears false. The maxims of action that apply to the actors involved in such cases appear to make it right for one actor to prevent what it is right for another actor to do.

Yet deontologists need not, and many would not, embrace such a conclusion. They might plausibly maintain that those who save their own (innocent) lives at the cost of other innocent lives are perhaps excused, but they are not justified.¹⁶³ Morality does not contain a selfish tipping principle that allows persons to prefer their own innocent lives to the innocent lives of others.¹⁶⁴ Rather, all individuals are bound by an agent-relative obligation not to take an innocent life. Insofar as the actors who confront one another in these cases are equally innocent, each is obligated not to sacrifice the life of the other to save his or her own life. While this may seem to the actors involved to require them to be heroes in circumstances in which they find themselves incapable of heroism, this is a reason merely to excuse them for

and that that promise constitutes an agent-relative obligation to do so that conflicts with the citizen's agent-relative permission to break the law. On this analysis, whether a judge has the power to promise punishment of the justified depends on, and thus cannot be used to establish, the morality of punishing the justified. See *infra* text accompanying notes 176-77.

163. I owe the articulation of this response to Stephen Morse.

164. Rejection of a selfish tipping principle follows from the assumption of what Nagel has described as "the impersonal standpoint" — the standpoint from which each actor must recognize that "[e]veryone's life matters as much as his does, and his matters no more than anyone else's." NAGEL, *EQUALITY AND PARTIALITY*, *supra* note 12, at 14; see also NAGEL, *THE VIEW FROM NOWHERE*, *supra* note 12, at 152-53, 159-63.

the innocent deaths that they cause; it is not a reason to think them justified in causing those deaths.

This rejoinder preserves the correspondence thesis by making it wrong for innocent actors to preserve their own lives at the cost of other innocent lives. In the event that an innocent actor violates his obligation not to use deadly force against another innocent actor, he loses his innocence. It then becomes right for the other to defend herself, for in so doing she is not violating the maxim against taking the life of an innocent person. Thus, neither gladiator may initiate combat against the other, but if one does, the other acts rightly in defending himself. And neither a police officer nor an innocent shield may initiate the use of deadly force against the other, but if one does, the other acts rightly in defending herself. The rightness of one actor's conduct (in refusing to use deadly force against an innocent person) thus determines the rightness of the other's permission of that conduct (that is, the rightness of not initiating the use of deadly force). And the wrongness of one actor's conduct (in initiating deadly force) determines the rightness of the other's prevention of that conduct (that is, the rightness of using deadly force in self-defense).

It thus appears that deontologists need not give up the correspondence thesis when confronted with genuine cases of gladiatorial confrontation. They need not think that gladiatorial maxims apply to gladiators.

3. *Self-Defense*

Deontologists preserve the correspondence thesis in cases involving gladiators and human shields by construing such cases as traditional cases of self-defense: both actors are wrong in such cases to initiate combat, but, in the event that one actor does the wrong thing, the other does the right thing by employing deadly force in self-defense. Some, however, may doubt that a proper deontological account of self-defense reflects the correspondence thesis. They may argue that, even in the classic case of self-defense, in which an innocent victim confronts a culpable aggressor, right action is defined by compliance with competing maxims. In such cases, actors rightly defend themselves only in the sense that they do what they are permitted to do. While the correspondence thesis may be true as a thesis about obligations, it is false as a thesis about permissions. Hence, while an actor may be permitted to defend herself, another actor may be permitted, or obligated, to prevent her from defending herself.¹⁶⁵ Let us

165. This argument was cleverly advanced during different colloquia discussions by George Christie, Samuel Freeman, and Jason Johnston, thus prompting the inclusion of this section.

examine three possible motivations for this position, together with what I take to be sound reasons for its rejection.

a. *The argument from competing permissions.* First, it may be tempting to think that, while obligations cannot conflict, permissions can. One individual may be permitted to do what another is permitted to prevent. But permissions are classically thought to constitute rights. If one has a permission to do an act, one has a right to do it.¹⁶⁶ If one has a right to do an act, then others have a duty not to interfere with one's performance of that act.¹⁶⁷ Hence, others cannot be permitted to prevent what one has a permission to do. To maintain both that permissions are rights and that permissions can conflict is thus incoherent. If another has a permission to interfere with one's performance of an act, then that actor has a right that one not interfere with her interference. But, if she has a right that one not interfere with that act of interference, then one has a duty not to interfere with that act. One thus cannot have a right to do the act with which she is permitted to interfere, for one cannot have both a right to do an act and a duty not to prevent others from preventing one's performance of that act.

b. *The argument from Hohfeldian privileges.* The second motivation for abandoning the correspondence thesis as a thesis about permissions might derive from a willingness to abandon the claim that permissions are best thought of as rights. If, instead, permissions are thought of as Hohfeldian privileges or liberties, then permissions could seemingly conflict. For if one has a Hohfeldian privilege to do an act, it is not the case that others have a duty not to interfere with that act; it is merely the case that others have no right that one not do that act.¹⁶⁸ But if others simply have no right that one not do the act, they may still have a Hohfeldian privilege to prevent one from doing the act.¹⁶⁹ Yet if they have such a liberty, then one does not have a right

166. In this section I am taking permissions to be distinct from what Wesley Hohfeld called *privileges* or *liberties*. Hohfeld maintained that it is important to keep "the conception of a right (or claim) and the conception of a privilege quite distinct . . ." WESLEY N. HOHFELD, *Some Fundamental Legal Conceptions As Applied in Judicial Reasoning*, in *FUNDAMENTAL LEGAL CONCEPTIONS AS APPLIED IN JUDICIAL REASONING AND OTHER LEGAL ESSAYS* 23, 39 (Walter W. Cook ed., 1919). According to Hohfeld, when one has a privilege to do an act it is not the case that one has a right to do the act (in the sense that others have a duty not to interfere with one's doing the act); it is rather the case that others have no right that one not do the act. *Id.* at 35-50. I discuss the understanding of permissions as Hohfeldian privileges in the next section. In this section, I am concerned with the more common understanding of permissions as rights. Under this conception, permissions count as a combination of both Hohfeldian claim rights and Hohfeldian privileges. *See id.* at 39.

167. *See id.* at 36-38.

168. *See id.* at 38-39.

169. *See id.* at 41.

that they not prevent one from doing the act. They thus have no right that one not do the act, but one has no right that they not prevent one from doing the act. The correspondence thesis appears to be in jeopardy, for the liberty that licenses one's act does not imply an absence of a liberty on the part of another actor to prevent that act.

This argument effectively admits that the correspondence thesis is true of obligations and permissions (as those are commonly understood) but denies that it is true of privileges or liberties. There are two lexically ordered points to be made in response to this argument. First, a deontologist can easily admit the argument. If there are liberties of the Hohfeldian sort, they define arenas of amoral action. Actors within such arenas are not bound by any maxims of action — they are genuinely at liberty. It is not the case that they are obligated to act in certain ways, but it is also not the case that they have rights (or commonly understood permissions) to act in certain ways (for on this argument, a liberty is not a right). Hence, actors operating under Hohfeldian liberties are untouched by deontological norms. That their actions may conflict is therefore of no normative importance, for their actions are of no normative importance. They are the actions of those in a state of nature. Because the correspondence thesis is a thesis about the conditions of moral action, it is neither troubling nor even surprising that it is inapplicable to amoral action.

Second, insofar as Hohfeldian liberties define amoral actions, they do not appear to capture the nature of acts done in self-defense or in violation of the law. As a matter of substantive morality, such acts appear to be of moral significance. A plausible moral theory may obligate an actor to employ self-defense, permit an actor to employ self-defense, or obligate an actor not to employ self-defense, but it at least speaks to the question of self-defense. Similarly, a plausible moral theory may obligate an actor to disobey the law, permit an actor to disobey the law, or obligate an actor not to disobey the law, but it at least speaks to the question of disobedience. Thus, while the correspondence thesis may be inapplicable in cases involving amoral acts (acts licensed by liberties), this does not make it inapplicable in cases involving self-defense or disobedience, for such acts are not amoral.

c. The argument from permissions to do wrong. The third motivation for rejecting the correspondence thesis as a thesis about permissions derives from the view that deontological norms of morality are exceptionless maxims: for example, "Do not kill." Justifications for violating such maxims are agent-relative permissions to do the wrong thing: for example, "You are permitted to kill in circumstances of self-defense." Such permissions do not remove the wrongness of vio-

lating moral maxims; they only entitle one to do what it is wrong to do (while making it virtuous to do what it is right to do, namely, to abide by the maxims in question). On this view, all permissions function like the permission that many theorists accord women concerning abortion: women's liberty interests are sufficiently great to permit them to make wrong decisions about abortion (for example, to have periodic abortions rather than to use birth-control).¹⁷⁰

According to this understanding of the content of deontological morality, an actor may be permitted to do the wrong thing while others are at least permitted, if not obligated, to do the right thing. Insofar as it is wrong for an actor to do what he is permitted to do, it must be right for others to prevent him from doing that act. Thus, an innocent actor whose life is in peril is permitted to kill a culpable aggressor in self-defense; but insofar as this act is wrong, albeit permitted, others are permitted, or perhaps obligated, to prevent it. Hence, the correspondence thesis is false: the rightness of an act does not determine the rightness of not preventing that act, for an act may be right only in the sense of being a permitted wrong, and it can be right to prevent a permitted wrong.

Deontologists might be particularly tempted to advance this analysis because of the payoffs that it might seem to offer in cases involving the punishment of citizens who disobey immoral laws. This analysis suggests the following. Citizens are subject to the exceptionless maxim, "Obey the law." In cases in which the law conflicts with other deontological maxims, citizens are permitted to disobey the law. But such a permission does not remove the wrongness of their disobedience. Insofar as such disobedience remains wrong, judges remain obligated to punish it. Hence, the correspondence thesis is false in cases of punishment, for the rightness of a disobedient act does not determine the rightness of not punishing that act. A disobedient act may be right only in the sense of being a permitted wrong, and, as was previously suggested, it can be right to punish a permitted wrong.

There are two responses to those who invoke this argument to defend the morality of preventing or punishing permitted acts. First, the argument fails because it is committed to the incoherent conclusion that permissions are rights, but that those to whom they apply are simultaneously under duties not to interfere with the attempts of others to prevent the exercise of those rights.¹⁷¹ If an actor has a right — even a right to do wrong — then others have a duty not to interfere

170. See, e.g., JUDITH J. THOMSON, *A Defense of Abortion*, in *RIGHTS, RESTITUTION AND RISK* 1-191 (William Parent ed., 1986).

171. See *supra* text accompanying notes 166-67.

with the actor's exercise of that right. Thus, even if permissions simply license wrong acts, rather than make otherwise wrong acts right, they generate duties on the part of others that are inconsistent with having rights of interference. Thus, the correspondence thesis in its original form remains intact: the permission of an act entails the rightness of not preventing that act.

Second, while autonomy rights, such as those of women concerning reproductive decisions, may be plausibly characterized as agent-relative permissions to do wrong, the justification of self-defense is quite implausibly described as an agent-relative permission to do wrong. Defending oneself against a culpable aggressor is not the wrong thing to do (albeit a permitted thing to do); it is, rather, the right thing to do. Thus, acts that prevent self-defense are not acts that prevent wrong action; they are acts that prevent right action. Even if critics could escape the conclusion that the correspondence thesis applies even to permissions to do wrong acts, they could not escape the correspondence thesis in cases of self-defense.

The same applies in cases of justified disobedience of the law. In such cases, the law conflicts with certain agent-relative obligations and permissions. Disobedience cannot constitute a permitted wrong in such cases; indeed, it cannot constitute a wrong at all. For such a conclusion would commit the deontologist not just to maxims that are gladiatorial, but to maxims that are genuinely contradictory. An actor would be simultaneously obligated both to obey the law and to disobey the law. If the content of deontology is at least constrained by the requirement that it not be contradictory, then a deontological theory cannot both impose an exceptionless obligation to obey the law and require citizens sometimes to break the law. Insofar as it obligates citizens to break the law in particular instances, a citizen's fulfillment of such an obligation must thus constitute the right thing to do; it cannot simply serve as a permitted wrong thing to do. Hence, a judge's punishment of such an act cannot be the punishment of a wrong (albeit permitted) act; it must be the punishment of a right act. And if the remaining cases convince us that deontologists should consider cases of punishment analogous to cases of prevention, then, even if critics could escape the correspondence thesis in cases that involve permissions to do wrong, they could not escape it in cases that raise the punishment of the justifiably disobedient.

4. *Status-Based Relationships and Contractual Roles*

Rather than stymieing the ability of deontologists to defend the correspondence thesis, the previous examples have illustrated the via-

bility of that thesis for a deontological morality. But this happy conclusion meets its most compelling counterexamples in cases such as the following. Imagine two mothers, each of whom has a small baby and neither of whom can swim. They find themselves aboard a sinking ship that has only one life vest. If each mother cannot secure that life vest for her own baby, her baby will surely drown. Is it not the case that each mother does the right thing in attempting to obtain the life vest for her own child, even when this amounts to preventing the other from saving her child?¹⁷² If so, the correspondence thesis is false, for the rightness of one mother's actions does not determine the rightness of the other mother's permission of those actions.

This sort of case tests the suggestion that individuals are entitled to prefer the welfare of those who are close to them to the welfare of strangers. Such a claim is most at home within a deontological morality. In light of the consequentialist requirement that all persons' interests be equally considered (such that each counts for one and only one), consequentialists have difficulty justifying why one might be morally justified in aiding a loved one in circumstances in which the interests of strangers are of equal or greater weight. Deontologists, on the other hand, would seem to be in a position to maintain that actors have agent-relative permissions or obligations to favor the interests of loved ones over the interests of strangers. But if the most plausible deontological morality contains such agent-relative permissions or obligations, the correspondence thesis is in jeopardy. For morality might then pit us against one another in gladiatorial moral combat.

It is possible to distinguish at least five deontological views concerning the content of the maxims that apply to actors in circumstances of this sort. Some of these views falsify the correspondence thesis, while others preserve it. Let us work through these to determine how likely it is that deontologists will consider themselves compelled to abandon the correspondence thesis in order to vindicate the intuitive claim that we may accord preferential treatment to those we love. Throughout this discussion we should not lose sight of our ultimate goal: to determine whether the correspondence thesis is implied by our most plausible deontological theory, so as to assess whether or not our most plausible deontological theory would license the punishment of the justified.

a. The argument from agent-relative obligations/permissions of preference. On the first view, actors are obligated, or at least permit-

172. This example is drawn from NAGEL, EQUALITY AND PARTIALITY, *supra* note 12, at 172.

ted, to give aid to those with whom they share special relationships as opposed to those with whom they share no relationships. Thus, when a choice must be made, parents are obligated (or permitted) to aid their children, friends are obligated (or permitted) to aid their friends, and so forth. This view flatly contradicts the correspondence thesis because it makes agent-relative obligations and permissions role-relative. The role of a mother is in part defined by the obligation (or permission) to administer preferential aid to her own children. The role of a lawyer is in part defined by the obligation (or permission) to provide special services to his clients.¹⁷³ Insofar as role-relative obligations and permissions can conflict, the correspondence thesis is false as a thesis about the conditions under which one actor must permit the actions of another. Just as in the case of the two mothers who must compete for one life jacket, it may be right for an actor in one role to prevent what it is right for an actor in another role to do.

Deontologists who are persuaded that certain roles create agent-relative obligations (or permissions) to administer preferential treatment may think that they are committed to a slippery slope that forces them to admit the possibility that other roles will create other sorts of agent-relative obligations and permissions. If mothers have agent-relative obligations to aid their children, then police officers may have agent-relative obligations to apprehend criminals, and judges may have agent-relative obligations to punish all citizens who violate the law (even those who do so justifiably).¹⁷⁴ Yet deontologists need not slide down such a slope.

Deontologists must recognize at least a *prima facie* distinction between two different sorts of roles: those that arise by virtue of a person's status (for example, the status of being a mother or a brother) and those that arise by virtue of a person's promise or contract (for example, a promise to provide security or a contract to perform legal services). Deontologists may plausibly believe that the only reason that contractual roles create certain agent-relative obligations of preferential aid is because these obligations are part of what persons have contracted for in taking on such roles. The agent-relative obligations that define such roles are thus special cases of the agent-relative obligation to keep promises. Thus, the agent-relative obligations that define contractual roles are given by the contracts in question.¹⁷⁵ The

173. For a defense of the claim that citizenship constitutes a unique role characterized by unique rights and duties, see Leslie Green, *Law, Legitimacy and Consent*, 62 S. CAL. L. REV. 795, 818-25 (1989); see also Hurd, *supra* note 22, at 1660 n.82 (discussing Green's role-relative conception of the duties of citizenship).

174. See NAGEL, *supra* note 12, at 151.

175. But see Green, *supra* note 173, at 818-25, for a defense of the claim that voluntarily

reason that a lawyer has a duty to provide preferential treatment to her clients is because she entered into a contract to do so. The reason that a husband has an obligation to provide aid to his wife is because he promised to do so as part of his marriage oath. Similarly, the reason that roles create other sorts of obligations (besides obligations of aid) is because those obligations have been voluntarily undertaken. Thus, the reason that a police officer is obligated to apprehend criminals is because he entered into an employment contract to do so, and the reason that judges must punish all individuals who violate the law, including those who do so justifiably, is because they took an oath of office that is rightly interpreted to include a promise to do so.

But if deontologists impose agent-relative obligations on certain actors just because actors have promised to fulfill those obligations, then deontologists must have a theory that specifies the conditions under which individuals have the power to contract into certain obligations. Just as in the case of consent discussed earlier,¹⁷⁶ individuals should be accorded the power to promise to do certain actions only if it is normatively justifiable for them to do so. On pain of circularity, the fact that an actor has made a promise cannot be a reason to maintain that he should fulfill that promise. We must have some reason, independent of his promise, to think that his promise is binding and, hence, that it entails certain role-relative obligations. As in the case of consent, there are reasons to think that certain promises make possible morally valuable institutions and relations. But the bindingness of a promise, like the scope of an actor's consent, depends upon the morality of the institutions and relations that it makes possible.

We have reasons to think that a promise to provide security services makes possible morally valuable relations. But we also have reasons to think that such a promise loses its normative power when it is made by a contract killer who takes security services to include the execution of an employer's political enemies. We have reasons to think that a promise to provide legal defense services makes possible certain morally valuable relations. But we also have reasons to think that defense lawyers properly lack the power to promise to destroy incriminating evidence. We have reasons to think that a judge's oath of office makes possible certain morally valuable institutions. But do we have reasons to think that that oath can include a binding promise

adopted roles can, and typically do, carry nonvoluntary duties — that is, duties that are not themselves subject to individual alteration. See also Serena Stier, *Legal Ethics: The Integrity Thesis*, 52 OHIO ST. L.J. 551 (1991), for an excellent discussion of the nonvoluntary duties imposed on lawyers by the value of integrity.

176. See *supra* text accompanying notes 160-62.

to punish the justified? This is, once again, the question we are trying to answer.

Just as deontologists could not use the possibility that justifiably disobedient persons would consent to their own punishment as a reason to think that their punishment is consistent with a plausible deontology, so deontologists cannot use the judicial oath of office to establish that judges have agent-relative obligations to punish the justified. Only if deontologists have *other* reasons to think that the punishment of the justified is itself justified will they have good cause to consider judges obligated to punish the justified if their oath of office includes such a promise.¹⁷⁷

This discussion makes clear that deontologists must be discriminating with respect to the sorts of agent-relative obligations that attach to particular roles. In cases in which deontologists ascribe agent-relative obligations to actors solely because those actors have assumed those obligations by promise or contract, they must be clear that those agent-relative obligations are not obligations just because they have been assumed by promise or contract. If, as I suspect, the conditions that justify contractually based agent-relative obligations embody the correspondence thesis, then actors may not promise to do acts that prevent other actors from doing what it is right for them to do.

Thus, a plausible deontological theory might contain status-based agent-relative obligations that violate the correspondence thesis without containing contractually based agent-relative obligations that violate the correspondence thesis. Recognizing this allows deontologists to have what they may plausibly consider the best of both worlds. They can vindicate our deep-seated conviction that the role of a mother includes the obligation to prefer the welfare of her own children to the welfare of other children (because they can insist that deontological morality contains certain status-based agent-relative obligations that violate the correspondence thesis). But they can deny that this commits them to a morality that empowers judges to assume contractual obligations to punish the justified (because they can deny that deontological morality contains contractually based agent-relative obligations that violate the correspondence thesis).

b. The argument from the denial of agent-relative obligations/permissions of preference. Some deontologists may find the above analysis altogether unsatisfying. First, they may be unhappy with the implications of the analysis in the case of the two mothers aboard the sinking ship, because it admits that, in such a case, both mothers confront

177. See *supra* text accompanying notes 160-62.

each other as moral gladiators. Each does right to attempt to save her own child, even though this means that the other will fail and her child will drown. Second, they may be unhappy with the suggestion that they can suspend the correspondence thesis in such cases without suspending it in cases that present judges with a decision about whether to punish the justified. They may subscribe to one of two claims: (1) that actors in contractually based roles also have agent-relative obligations (or permissions) that are status-based, so that actors in such roles may be right to prevent what it is right for others to do;¹⁷⁸ or (2) that the theory that grounds gladiatorial status-based obligations will speak in favor of a theory that allows individuals to create by contract gladiatorial obligations that make it right to prevent others from doing what it is right for them to do.

To alleviate these concerns, deontologists must reject the claim that there can be any agent-relative obligations or permissions that violate the correspondence thesis. They must deny that mothers have obligations or permissions to aid their own children when doing so will prevent others from aiding other children. Such a denial might well be deontologically defensible. The discussion concerning gladiators and human shields made plausible the view that morality does not contain a selfish tipping principle that makes it right to prefer one's own innocent life to the innocent life of another.¹⁷⁹ The same might be said of cases in which persons prefer the lives of their loved ones to the lives of strangers: neither mother on the sinking ship is right to prefer the life of her own child to the life of the other mother's child. If one of these mothers finds this conclusion to be too difficult to bear, and so seizes the life jacket for her own child, we may find grounds for thinking her excused, but she is not justified. Like the gladiator who takes up arms or the police officer who shoots through an innocent shield, the mother who seizes the life jacket may act understandably, but she does not act rightly.

If this analysis of cases in which persons are tempted to provide preferential treatment to loved ones is the better of the two here considered, then we have grounds to think that the correspondence thesis is secure as a thesis about preventative acts. This provides us with a further reason to suppose that, as a general matter, roles (of either the status-based or contractually based sort) cannot call upon those who occupy them to prevent others from doing what it is right for them to do. We thus have grounds upon which to maintain that official oaths

178. See *supra* note 175 and accompanying text.

179. See *supra* notes 163-64 and accompanying text.

of office cannot include promises to punish the justified that generate agent-relative permissions or obligations to do so.

Deontologists who remain undecided about which of the above accounts to embrace may continue to seek middle ground from which to defend both the claim that morality does not provide role-relative permissions or obligations to provide preferential aid to loved ones and the claim that morality does not prohibit preferential treatment in all cases. Three compromise positions present themselves.

c. *The argument from Hohfeldian liberties of preference.* The first such compromise stems from the claim that morality runs out in circumstances in which we find ourselves forced to choose between aiding a loved one and aiding a stranger. In such circumstances, we are genuinely at liberty to give preferential treatment to our loved ones, for while we have neither an obligation nor a permission to prefer our loved ones to strangers, strangers have no right that we not do so. To maintain this, one must maintain that such circumstances of third-party defense are amoral. This was hard to believe in the case of self-defense, and it seems harder to believe in these cases. However, as I argued previously, if these cases of third-party defense suspend the correspondence thesis by constituting instances in which agents possess Hohfeldian liberties, then they do not pose a reason to reject the correspondence thesis in cases of punishment.¹⁸⁰ While these cases may stand beyond the reach of moral maxims, cases of punishing others do not.

d. *The argument from preferential killings versus preferential omissions to save.* A second compromise might be struck by deontologists who similarly admit that there are no obligations or permissions to prefer the welfare of loved ones to the welfare of strangers, but who also argue that, while persons have agent-relative obligations not to kill, they do not have agent-relative obligations to save.¹⁸¹ In the case of the two mothers aboard the sinking ship, this position would imply that neither mother has a right to seize the life jacket if it is already in the possession of the other. For to do so would *cause* the death of that

180. See *supra* text accompanying notes 168-69.

181. For a defense of this position, see FINNIS, *supra* note 11 at 118-25, 195; MICHAEL MOORE, ACT AND CRIME (forthcoming 1992); Epstein, *supra* note 56, at 198-200; Richard A. Epstein, *Causation and Corrective Justice, A Reply to Two Critics*, 8 J. LEGAL STUD. 477 (1979); Michael S. Moore, *Torture and the Balance of Evils*, 23 ISRAEL L. REV. 280, 299-300, 308-12 (1989). For critiques of this view, see SINGER, *supra* note 5, at 149-53; PHILLIPA FOOT, *The Problem of Abortion and the Doctrine of Double Effect* in VIRTUES AND VICES AND OTHER ESSAYS IN MORAL PHILOSOPHY 19, 25-27 (1978); Phillipa Foot, *Morality, Action and Outcome*, in MORALITY AND OBJECTIVITY 23, 24 (Ted Honderich ed., 1985); Richard A. Posner, *Epstein's Tort Theory: A Critique*, 8 J. LEGAL STUD. 457, 460 (1979); Ernest J. Weinrib, *The Case for a Duty to Rescue*, 90 YALE L.J. 247 (1980).

mother's child and so violate the obligation not to kill. But neither mother has an obligation to give up the life jacket if it is already in her possession, for to refuse to transfer the life jacket from her own child to the other mother's child would simply be to omit to save that child, it would not be to kill that child.

This thesis entails that if neither mother has previous possession of the life jacket, then neither mother may take the life jacket at the time of crisis. To do so would kill the other mother's child by affirmatively causing its death. Hence, if the life jacket is not already in the possession of one of the two mothers, both mothers do the right thing only if each refuses to grab the jacket. While both children will die, they will not have been killed; they will simply not have been saved.

This position preserves the correspondence thesis, for the maxims that apply to each mother obligate the other not to interfere with their fulfillment. If one mother possesses the life jacket, the other does right not to take it from her. If neither mother possesses the life jacket, each does right only if neither takes it. Thus, on this account of the content of morality, one actor is never justified in preventing what the other actor is justified in doing. If this constitutes the best analysis of these sorts of cases, then this analysis, like the second one advanced above, provides a basis for thinking that the correspondence thesis should be preserved in cases of punishment.

This position constitutes a compromise between the first two views with which we began, for while it denies that there are obligations and permissions of preference, it makes room for circumstances in which persons can prefer the welfare of their loved ones to the welfare of strangers, namely, when their aid to loved ones does not cause harm to strangers (though it may constitute a failure to save strangers from harm). But this compromise commits its defenders to maintaining that, in the original case of the two mothers aboard the sinking ship, both of their children should be sacrificed, when one can be saved. While deontologists cannot permit consequential calculations to dictate the content of their morality, they may nevertheless find that a plausible deontological theory would not contain maxims that commit them to such a troubling conclusion. They thus may search further for a compromise.

e. The argument from the intent with which preferential treatment is bestowed. The third compromise follows the previous two in admitting that there are no obligations or permissions that make it right to prefer loved ones to strangers. It also incorporates the claim that individuals are obligated not to kill but are not obligated to save. It denies, however, that actors violate the obligation not to kill in

circumstances in which they cause death without a direct intent to do so.¹⁸² If applied to our previous hypothetical, each mother may save her own child by seizing the life jacket from the other, or from some neutral resting place, for by so doing she does not directly intend the death of the other mother's child; she simply knows that that death will come about if she is successful.

This view constitutes a compromise because, like the previous position, it denies the existence of obligations or permissions that make it right to prefer the welfare of loved ones to that of strangers while carving out circumstances in which the exercise of such a preference does not violate any agent-relative maxims. By licensing the rescue of one life when both cannot be saved, this view also escapes the troubling conclusion to which the previous compromise position was committed. If plausible, this position implies that we must abandon the correspondence thesis in cases like that of the two mothers aboard the sinking ship because, in such cases, one actor does not do wrong in preventing the other from doing what it is right for her to do.

Yet this argument is guilty of the same confusion that beset previous consequentialist attempts to limit rightmaking consequences on epistemic grounds: it smuggles considerations of culpability into the specification of the conditions of right action. Actors who *knowingly* violate moral maxims may indeed be less culpable than actors who intentionally violate moral maxims, but before we can engage in this discussion we must have some independent understanding of the content of moral maxims — some theory of right action over which intentions can range. If it is wrong to kill one innocent child to save another, then it is wrong to do so negligently, recklessly, knowingly, or intentionally — though it may not be as culpable to do so negligently as opposed to knowingly, or knowingly as opposed to intentionally. It is even wrong to do so innocently — that is, on a reasonable belief that it is not wrong, though such a wrong act is wholly nonculpable.

Thus deontologists cannot avail themselves of this compromise without confusing the metaphysics of right action with the conditions of culpability. As this compromise violates the correspondence thesis precisely because it commits such a confusion, it poses no threat to the viability of the correspondence thesis for a plausible deontological morality.

Where does this leave us? We have worked through a series of cases that on their face appear to function as counterexamples to the

182. The Catholic Church takes this position. See H.L.A. HART, PUNISHMENT AND RESPONSIBILITY 122-25 (1968).

claim that a viable deontological moral theory would embody the correspondence thesis. We have found that with regard to all of these cases deontologists can plausibly adopt one of two responses: they can deny that the maxims properly applicable to actors in these cases violate the correspondence thesis by being gladiatorial in content, or they can admit that the maxims properly applicable to actors in these cases violate the correspondence thesis but deny that such a violation serves as a precedent for suspending the thesis in cases involving punishment. If deontologists can deliver the substantive moral arguments required to ground these responses adequately, then we have reason to suppose that a plausible deontological theory embraces the correspondence thesis — at least in cases that serve as analogues to the cases that raise the question of whether or not to punish the justified.

When coupled with the general conclusion reached in the previous section, this result prompts the conclusion that as a matter of our best moral theory — be it consequentialist or deontological — the punishment of the justified cannot be justified. We must seek to resolve the dilemma with which we began our lengthy discussion by means other than abandoning the punishment principle. That is, if the punishment principle is in genuine conflict with the rule of law in cases that call upon judges to adjudicate justified acts of private disobedience, then judges must sacrifice the rule of law. And if the punishment principle is in genuine conflict with the separation of powers in cases that call upon system designers to adjudicate justified acts of judicial disobedience, then system designers must sacrifice the separation of powers. For morality does not permit the sacrifice of the punishment principle.

CONCLUSION

I have argued that our traditional understanding of the jurisprudential and political requirements provided by the rule of law and the separation of powers is incompatible with the demands of morality. According to our traditional conception, the rule of law and the separation of powers combine to require judicial deference to legislatively enacted rules in cases in which such deference cannot be morally justified. These principles call upon judges to punish citizens who are morally justified in breaking the law, and they call upon institution designers to punish judges who are morally justified in acquitting justifiably disobedient citizens. Our understanding of the rule of law and the separation of powers therefore violates what I have called the punishment principle — the moral prohibition against punishing actors who act morally.

The dilemma generated by the incompatibility of our fundamental

legal, political, and moral principles has been thought by many to be superficial. Some have sought to resolve it by defending the claim that morality invests the law with practical authority. Such authority makes any act of disobedience, private or judicial, morally unjustifiable. Hence, judges never punish the justified when punishing the disobedient. I have argued both here and in my past work that this solution is no solution at all, because the concept of practical authority required to make this a solution is conceptually indefensible. Thus the authority of the law cannot make obedience moral in cases in which the law departs from the demands of morality.

Those who have concurred with my view have been compelled to admit that so long as the law departs from morality in particular cases, individuals will be genuinely justified in departing from the law. They have thus been forced to solve the dilemma engendered by our competing principles by abandoning one or more of those principles. Their principle of choice has been the punishment principle. In rejecting this principle, theorists have sought to demonstrate that morality licenses the punishment of the justified. In many cases, they argue, morality makes it right to punish citizens who are right to disobey the law. In those cases in which morality does not make it right to punish citizens who are right to disobey the law, it nevertheless makes it right to punish judges who refuse to punish such citizens.

Those who have sought to defend this paradoxical thesis have maintained that morality is inherently role-relative. Institutional roles generate unique reasons for action — reasons that sometimes make it right for actors within those roles to do what actors outside of those roles are wrong to do. The reasons for action given by the values inherent in the rule of law are unique to those who assume judicial roles. In many cases, these role-relative reasons are sufficiently powerful to justify judges in punishing citizens who rightly disobey the law in disregard of these reasons for obedience. The reasons given by the values inherent in democracy and the separation of powers are unique to those who perform the systemic task of maintaining structural pluralism. In cases in which the rule-of-law values are insufficient to justify a judge in punishing a justifiably disobedient citizen, these role-relative reasons are sufficiently powerful to justify the punishment of that judge. Insofar as institutional actors have role-relative reasons that justify the punishment of the justified, the punishment principle must be abandoned. The dilemma is solved, for our jurisprudential and political principles do not demand anything that morality prohibits.

This solution has the virtue of justifying the legal and political status quo. It does not call upon us to rethink our commitment to the

separation of legislative and adjudicative powers, and it does not demand that we give up the sort of rule-governed adjudication that that separation, coupled with the constraints imposed by the rule of law, appears to require. It vindicates those who espouse the virtues of judicial restraint and provides a basis for criticizing those who advocate judicial activism.

But it also punishes the justified. And as I have argued herein, an appeal to role-relative morality cannot justify the punishment of the justified. For morality is not role-relative. On our best consequentialist account of morality, all consequences causally generated by an act are reasons for action that enter into the balance that determines the rightness of that act. If a citizen's disobedient act causes judges and system designers to render decisions that affect the rule of law and the separation of powers, then the values inherent in the rule of law and the separation of powers are reasons for action that must be included in the balance that determines the rightness of the citizen's disobedience. If that balance nevertheless favors disobedience, the citizen's act is right, and the citizen's subsequent punishment is then wrong, as is the subsequent punishment of the judge who rightly acquits that citizen. On our best deontological account of morality, certain roles may provide actors with agent-relative obligations. But the institutional roles occupied by judges and system designers can provide such obligations only by virtue of being contractually created (i.e., by oaths of office). And our best understanding of the content of deontological norms suggests that institutional actors cannot contractually assume an obligation to punish the justified. Hence, if a citizen's disobedience is right, it cannot be right to punish that citizen, and it cannot be right to punish a judge for not punishing that citizen.

We must thus rethink our traditional understanding of the rule of law and the separation of powers. Our task is to square our commitment to those principles with the recognition that their value cannot lie in their enforcement of rule-governed adjudication when the rules that are said to govern adjudication depart from the rules of morality. This is a task that calls upon us to vindicate the significance of imperfect positive law while recognizing that the only law that can be *legally* binding is natural law. Positive law does not possess the authority to give private and official actors reasons to do anything other than what is morally required of them, and morality does not justify the punishment of those who in fact do what is morally required of them.

I have elsewhere begun this task of reconceptualizing the importance of our systemic values with my defense of why democratic rules could and would continue to guide private and official conduct even if

they were fully recognized to possess only theoretical authority.¹⁸³ This defense draws upon the epistemic significance of legislatively enacted rules, and the values inherent in their systemic application, to the continued project of constructing and implementing our best understanding of the requirements of morality. If we have reason to think that positive law is good evidence of natural law, then we have reason to think that private and official actors should in most cases adhere to positive law. But their obligation to do so can be, at most, an epistemic obligation; it cannot be either a moral or a legal one. Hence, in cases in which the law clearly departs from morality (understood as embodying the complex set of institutional values herein discussed), neither citizens nor judges can have any obligation to abide by it.

183. See Hurd, *supra* note 22, at 1675-77.