

Self-Defense: Agent-Neutral and Agent-Relative Accounts

Jeremy Waldron†

INTRODUCTION

One of the innumerable debts I owe Sandy Kadish is in respect of his warm and consistent encouragement to invest some portion of my philosophical energies in the problems of the criminal law. When I replied that what I knew most about was the history of political philosophy, Sandy was undeterred. “Kant wrote about criminal law,” he would say. “So did Hobbes and Locke. Those guys talked about capital punishment, individual responsibility and self-defense, just as much as they dealt with topics in political theory like the nature of law and the obligation to obey.” He was right. There is plenty to discuss and plenty of inspiration to be drawn from the canon of traditional political philosophy, so far as criminal jurisprudence is concerned.

So here is a response to my dear friend and mentor. In this Essay, I would like to consider the answers given by Thomas Hobbes, John Locke, and Sanford Kadish to the following question, which continues to exercise us in modern criminal jurisprudence:

If a person is to be exculpated—for example, in respect of the killing of another human being—because his act was justified, must it be justified from a point of view that transcends his own (socially justified, for example, or justified from a God’s-eye point of view), or is there room for a concept of justification that is, in philosophical terminology, agent-relative?¹

Copyright © 2000 California Law Review, Inc. California Law Review, Incorporated (CLR) is a California nonprofit corporation. CLR and the authors are solely responsible for the content of their publications.

† Maurice and Hilda Friedman Professor of Law and Director of the Center for Law and Philosophy, Columbia University.

1. For the notion of agent-relativity, see the following: THOMAS NAGEL, *THE VIEW FROM NOWHERE* 164-66 (1986); DEREK PARFIT, *REASONS AND PERSONS* 27 (1984); and SAMUEL SCHEFFLER, *THE REJECTION OF CONSEQUENTIALISM* 41-114 (1982). A moral theory is agent-relative (or “agent-centered”) if it assigns different agents different goals. Egoism is an example: *X* is to pursue the well-being of *X*, *Y* the well-being of *Y*, etc. A theory is agent-neutral if it assigns different agents the same goal. Utilitarianism is an example: *X* is to pursue the well-being of *X* and *Y* and *Z* . . . , *Y* is to pursue the well-being of *X* and *Y* and *Z* . . . , etc. A moral restriction (such as “Do not kill”) is agent-relative if

The consideration of Professor Kadish's work in this context will raise no eyebrows: He has written on exactly this issue, and he has done so in the best tradition of focused philosophical inquiry into the foundations of the criminal law. It is the inclusion of Hobbes and Locke in this distinguished company that requires explanation. Kadish, as I have said, will understand. But others may not. After all, neither Hobbes nor Locke is thought of primarily as a jurist. Their works are not routinely cited in law books or in litigation. They are generally taken to be more interested in political than in legal philosophy. Still, I think there are several reasons for considering what they have to say on this issue.

First, their writing is valuable precisely because they are not jurists. One of the hardest things to do in any discussion with lawyers about the moral principles underlying the law is to get them to stop regarding the doctrinal utterances of judges as data sufficient to refute any moral theory which those utterances may contradict. Hobbes and Locke write about self-defense for the most part untainted by judicial discussion; certainly their work is uncontaminated by the legalistic distinctions we have evolved in the common law.² They are not authorities, of course. But their theories provide a perspective from which we can critically reconsider all our doctrinal thinking on the topic, instead of proceeding as we do in legal argumentation, at most using some doctrines and precedents to criticize others.

each agent is to be concerned as a matter of priority with his own refraining from killing—for example, if he may not kill in order to prevent a greater number of prohibited killings by others. Robert Nozick's theory of rights in *ANARCHY, STATE, AND UTOPIA* (1974) is a theory of agent-centered restrictions along these lines. A moral permission is agent-relative if an agent is held entitled himself to give greater weight to some interest of his in his moral calculations than others are required to give to it (that is, to that interest of his) in theirs. Samuel Scheffler defends a theory of agent-relative permissions (he calls them "agent-centered prerogatives") as a modification of utilitarianism in *The Rejection of Consequentialism*. My thesis in this Essay is that we may want to consider self-defense as establishing something like an agent-relative prerogative.

2. Indeed Locke made a virtue of this. The law of nature, he said in the *Second Treatise*, is as intelligible and plain to a rational Creature, and a Studier of that Law, as the positive Laws of Common-wealths; nay, possibly plainer; As much as Reason is easier to be understood, than the Phansies and intricate Contrivances of Men, following contrary and hidden interests put into Words; For so truly are a great part of the Municipal Laws of Countries, which are only so far right, as they are founded on the Law of Nature, by which they are to be regulated and interpreted.

JOHN LOCKE, *TWO TREATISES OF GOVERNMENT* 275 (Peter Laslett ed., Cambridge Univ. Press 1988) (1690) (emphasis omitted) [hereinafter LOCKE, *TWO TREATISES*]. See also Locke's scathing comments about lawyerly reasoning in the *ESSAY CONCERNING HUMAN UNDERSTANDING* 490-98 (Peter H. Nidditch ed., Oxford Univ. Press 1975) (1690) [hereinafter LOCKE, *HUMAN UNDERSTANDING*]. There is a general discussion of Locke's views of law in JEREMY WALDRON, *THE DIGNITY OF LEGISLATION* 63, 76 (1999).

On the other hand, we must also bear in mind Thomas Hobbes's insistence that although "[t]he Law of nature . . . is a part of the Civill Law in all Common-wealths of the world," and although "Law can never be against Reason," still "[t]he Interpretation of the Lawes of Nature, in a Common-wealth, dependeth not on the books of Morall Philosophy." THOMAS HOBBS, *LEVIATHAN* 185, 186, 191 (Richard Tuck ed., Cambridge Univ. Press 1991) (1651).

Second, Hobbes and Locke discuss self-defense and self-preservation not as specialized topics in the philosophy of law but as pivotal issues in social and political theory. For Hobbes, self-preservation lies at the heart of the social contract; for Locke, self-defense lies at the heart of the theory of justified revolution. Both of them hold theories of self-preservation that have consequences running far beyond the criminal law, into more general issues about the nature and justification of political and economic systems. For this reason, they enable us to explore moral ramifications that might be left unconsidered in a purely legal context.

Third, the work of these natural law theorists is particularly apt for discussing the various legal issues that cluster around self-preservation. One or two of the leading cases³ mention the old idea that “necessity knows no law.” They suggest that the issues that arise are so extreme that we should think of the people involved as being in a state of nature, governed by the law of nature, rather than in a situation governed by positive law.⁴ Well, these ideas of the state of nature, natural law, and primeval overriding necessity are not just forensic rhetoric. They have been thought about in our culture, and if we take them seriously in this regard, we should look not only to their use in the courtroom, but to what has been said about them by those philosophers who have made that their business. Again, this is not because John Locke and Thomas Hobbes are authorities, but because, since these matters are fraught with difficulty on all accounts, any sophisticated attempt to think them through from first principles should command our attention.

3. See, e.g., *Kansas v. Mugler*, 29 Kan. 252 (1883) (due process constraints inappropriate for an emergency); *City of Buffalo v. Chadeayne*, 7 N.Y.S. 501, 503 (Sup. Ct. 1889) (due process may legitimately be abandoned in an emergency); *Phillips v. Hooker* 62 N.C. 193 (1867) (inhabitants of American South had no choice but to use Confederate currency during the Civil War).

4. The best-known example is found in the argument of counsel for the defense in the lifeboat case of *United States v. Holmes*, 26 F. Cas. 360, 366 (C.C.E.D. Pa. 1842) (No. 15,383):

We do not seek authorities for such doctrine. The instinct of these men's hearts is our authority—the best authority. Whoever opposes it must be wrong, for he opposes human nature. All the contemplated conditions, all the contemplated possibilities of the voyage, were ended. The parties, sailor and passenger, were in a new state. All persons on board the vessel became equal. All became their own lawgivers; for artificial distinctions cease to prevail when men are reduced to the equality of nature. Every man on board had the right to make law with his own right hand, and the law which did prevail on that awful night having been the law of necessity, and the law of nature too, it is the law which will be upheld by this court, to the liberation of this prisoner.

I

THE CASE OF THE INNOCENT THREAT

To sharpen our sense of the problem I want to consider, here is a hypothetical case. (I have adapted it slightly from one posed in Sanford Kadish's essay *Respect for Life and Regard for Rights*):⁵

A fanatical terrorist armed with various weapons has killed a number of people and is now holed up in a building surrounded by police. The police believe that if he is permitted to escape, he will kill many more, and they have decided they cannot allow this.

Unfortunately, the terrorist has a hostage, and he is determined to move out of the house using the hostage as a human shield. He knows that there is a good chance that the police will try to kill him even if this means shooting through the body of the innocent hostage. Devilishly, he explains the situation to the hostage and convinces him that either they will both escape together or they will both die together. He gives the hostage a pistol with which he can shoot back at any police officers who may try to shoot (through) him, assuring him of course that if he even so much as makes a move to turn the pistol on his captor he (the terrorist) will kill him instantly. The police, for their part, are fully aware that an escape attempt of this kind may be made.

The terrorist and his hostage begin moving down the stairway of the building. By sheer bad luck, an armed police officer has chosen just that moment to move to a position higher up that stairway. He turns a corner of the stair and sees the other two coming down towards him. All three individuals grasp the situation in a flash, and all three begin firing: the terrorist and his hostage firing at the officer, the officer firing at and through the hostage. Each of them is shooting to kill. None of them is under any misapprehension of fact.

In this story, we have a number of people shooting at each other, each trying to save his own skin. How are we to deploy ideas of justification, necessity, and self-defense in such a terrible case?

To begin with, presumably no one believes that the terrorist is justified in firing. He can make no valid claim of self-defense since he knows very well that if he were instantly to throw down his weapons, the other two would stop firing, and there would be no further danger to his life. But he doesn't, and in the circumstances his actions have the effect of making the hostage and the police officer mortal threats to one another, even though both of them are innocent. They fire on each other, both, as I have said, shooting to kill, each convinced that killing the other is the only way to save his own life. Are either or both of them justified in firing?

5. Sanford H. Kadish, *Respect for Life and Regard for Rights in the Criminal Law*, in *BLAME AND PUNISHMENT: ESSAYS IN THE CRIMINAL LAW* 109, 122-23 (1987).

Few will have difficulty saying the police officer is justified. In fact his action seems justified several times over. He is firing in self-defense to save his own life, and he has also either a general defense of necessity or a specific law enforcement defense justifying him in doing whatever it takes to arrest, disarm, or kill the terrorist.

The interesting question relates to the hostage. Does he have a valid claim of self-defense? It seems odd to say that he does not. He is an innocent party, and had no choice but to become a human shield for the terrorist. He is shooting only in order to save his own life, not to help the terrorist, and he would stop shooting in an instant if he thought the police officer was not shooting at or through him. He fires only because someone is trying to kill him, and this is the only way he can defend himself against attack.

But the consequences of saying that he *does* have a claim of self-defense are disconcerting. The provisions of most penal codes justify self-defense only in response to *unlawful* or *unjustified* aggression, and that is not the case here.⁶ We have seen that the police officer's shooting at the hostage is justified in self-defense (as well as on grounds of necessity). If the latter's shooting at the police officer is also justified in self-defense, we have the paradoxical result that each of two people is justified in trying to bring about the death of the other while keeping himself alive.⁷ That plays havoc with some traditional conceptions of justification. We say, usually,

6. The test may be interpreted liberally, so that *X* may defend himself against *Y*'s aggression even if (a) *Y*'s aggression is excused, for example by reason of *Y*'s insanity, or (b) *X* mistakenly believes that *Y*'s aggression is unjustified whereas it is in fact justified. But I have set up the case so that neither of these circumstances apply.

7. The paradox was sketched out by Robert Nozick in his discussion of what he called "innocent threats":

I shall not pursue here the details of a principle that prohibits physical aggression, except to note that it does not prohibit the use of force in defense against another party who is a threat, even though he is innocent and deserves no retribution. An *innocent threat* is someone who innocently is a causal agent in a process such that he would be an aggressor had he chosen to become such an agent. If someone picks up a third party and throws him at you down the bottom of a deep well, the third party is innocent and a threat; had he chosen to launch himself at you in that trajectory he would be an aggressor. Even though the falling person would survive his fall onto you, may you use your ray gun to disintegrate the falling body before it crushes and kills you? Libertarian prohibitions are usually formulated so as to forbid using violence on innocent persons. But innocent threats, I think, are another matter to which different principles must apply. Thus a full theory in this area also must formulate the *different* constraints on response to innocent threats. Further complications concern *innocent shields of threats*, those innocent persons who themselves are nonthreats but who are so situated that they will be damaged by the only means available for stopping the threat. . . . May one knowingly injure innocent shields? If one may attack an aggressor and injure an innocent shield, may the innocent shield fight back in self-defense (supposing that he cannot move against or fight the aggressor)? Do we get two persons battling each other in self-defense? Similarly, if you use force against an innocent threat to you, do you thereby become an innocent threat to him, so that he may now justifiably use additional force against you (supposing that he can do this, yet cannot prevent his original threateningness)? I tiptoe around these incredibly difficult issues here, merely noting that a view that says it makes nonaggression central must resolve them explicitly at some point.

Nozick, *supra* note 1, at 34-35 (citation omitted).

that to justify an act is to show that it leads to an outcome that is better (or less bad) than the situation which would result if it were not performed. But the outcome that the hostage is aiming to prevent (his own death at the hands of the police officer) is the very outcome the police officer is aiming to bring about. Since the first of these outcomes cannot be both better and worse than the second, it would seem that at most only one of the shootings can be justified.

Most discussions in the area that raise this possibility turn on some issue of mistake. In *Rethinking Criminal Law*,⁸ George Fletcher considers a case in which a shopkeeper fires on a police officer coming to his rescue, wrongly believing him to be one of the armed burglars who is attacking his shop; the police officer pinned down by the shopkeeper's fire has to ask himself whether he may shoot back in his own self-defense. But in the present case, no one is under any misapprehension; everyone knows all the facts about the situation and about the knowledge of each of the others. The question of justification that the case poses, then, is clean, uncontaminated by epistemic considerations. We cannot determine whether the hostage has a justification in my example without asking what sort of thing such a justification might be.

As we examine the views of our two canonical philosophers, Thomas Hobbes and John Locke, we will find that Hobbes's theory is quite clear that the hostage *does* have a justification, and it has no difficulty with the coexistence of that justification and the justification of the police officer's shooting to which the hostage's shooting is a response. Justification for Hobbes is agent-relative.⁹ The implications of Locke's account are less clear. The arguments with which he supports his views on self-defense indicate a theory of justification that is agent-neutral: An act is justified if it conduces to the preservation of mankind, as much as possible. But he combines this with a claim about the priority of self-preservation that seems more Hobbesian in its tendency, though (unlike Hobbes) Locke leaves us quite unsure of the grounds on which it is asserted. Locke's theory foreshadows the embarrassment of modern legal thinking on the matter. We want to have an objective theory of justification, but we are (as George Fletcher indicates)¹⁰ "drawn to" the idea that justifying an act may be a more personalized business in cases where, with the best will in the world, innocent people pose a threat to one another's survival.

8. GEORGE FLETCHER, *RETHINKING CRIMINAL LAW* 763 (1978).

9. For the definition of "agent-relative," see *supra* note 1.

10. Concerning the agent-relative approach, Fletcher writes: "It is difficult to know how much time to spend on this subjective and personalized theory of justification. So far as I know, no-one in the English-speaking world has defended the thesis in the literature. But it would not surprise me if many people were drawn to it." FLETCHER, *supra* note 8, at 767.

II

HOBBES ON SELF-DEFENSE

The most famous philosophical account of self-defense is surely that of Thomas Hobbes, in both *Leviathan*¹¹ and his earlier work *De Cive*.¹² In Chapter 21 of *Leviathan*, Hobbes puts forward the following exception to his principle that civil liberty is to be understood as nothing more than the silence of the laws:¹³

If the Sovereign command a man (though justly condemned,) to kill, wound, or mayme himselfe; or not to resist those that assault him; or to abstain from the use of food, ayre, medicine, or any other thing, without which he cannot live; yet hath that man the Liberty to disobey.¹⁴

Whereas in general "[t]he Liberty of a Subject, lyeth . . . only in those things, which in regulating their actions, the Sovereign hath praetermitted,"¹⁵ in this sort of case the subject retains liberty that is independent of and has priority over the sovereign's commands; and he has no obligation to obey.

Notice that the terms in which the Hobbesian doctrine is couched are wider than those we associate with self-defense. As well as the liberty "to resist those that assault him," there is also a liberty to disobey an order to harm oneself (for example, an order to execute as Socrates did a sentence of death on oneself), and the general liberty which we would associate with a very broad doctrine of necessity. Hobbes also goes on to connect self-defense with the privilege against self-incrimination,¹⁶ a right to resist conscription,¹⁷ and a right (comparable to that exercised by the good townspeople in *High Noon*) to refuse to join a posse.¹⁸ It is an attractive feature of Hobbes's approach that he provides argumentative unity for these disparate issues. Certainly we should not be seduced into thinking, by the fact

11. HOBBS, *supra* note 2.

12. THOMAS HOBBS, *DE CIVE* 58-59 (English version, Howard Warrender ed., Oxford Univ. Press 1983) (1651).

13. HOBBS, *supra* note 2, at 152-53.

14. *Id.* at 151.

15. *Id.* at 148.

16. *See id.* at 151 ("If a man be interrogated by the Sovereign, or his Authority, concerning a crime done by himselfe, he is not bound (without assurance of Pardon) to confesse it. . .").

17. *See id.* at 151-52:

Upon this ground, a man that is commanded as a Souldier to fight against the enemy, though his Sovereign have Right enough to punish his refusall with death, may neverthelesse in many cases refuse, without Injustice. . . . When Armies fight, there is on one side, or both, a running away; yet when they do it not out of trechery, but fear, they are not esteemed to do it unjustly, but dishonorably.

I will discuss the tension between subject's privilege and sovereign's right, and between injustice and dishonor, below. *See infra* text accompanying notes 38-39 and 61-66.

18. For a superb discussion, see JEAN HAMPTON, *HOBBES AND THE SOCIAL CONTRACT TRADITION* 132-88 (1986).

that lawyers distinguish rigorously, for example, between the justifications of self-defense and necessity, that the two must therefore have quite different grounds, or that a given case has to be one or the other.

Now, no one is interested in what Thomas Hobbes (or any other philosopher) says; we are interested in how he argues and what he can justify. What case does Hobbes make for this claim that the subject has this liberty to resist and disobey?

The argument ought to proceed in two phases:

- (1) Generally, in Hobbes's and in all contractarian philosophy, any liberty the subject has in civil society is a residue of the primeval liberty he enjoyed in the state of nature. This is particularly true of self-defense. So the first question to ask is why people have a natural liberty of self-defense.
- (2) We must ask why this liberty in particular is not one that a person is deemed to give up as he moves from the state of nature into civil society. What is it about this liberty that makes it peculiarly resistant to that transition, in Hobbes's view? What is the basis of its inalienability?

In fact, the best discussion of question 1 comes in Hobbes's attempt to address question 2 in *Leviathan* and *De Cive*. But as we shall see, the answer or answers he gives to question 2, if taken alone, would cast his theory of self-defense and necessity in a misleading light. Taken alone, at least one of the answers he gives to question 2 suggests that self-defense and necessity should be regarded as matters of excuse, not justification. It is only when it is connected with the independent answer that he gives to question 1 that we see the justificatory aspect of the matter reemerging.

I shall begin with question 2. When Hobbes talks about "liberty," he means something like "Hohfeldian privilege."¹⁹ To say that *P* is at liberty to do *A* is to say just that *P* has no obligation not to do *A*. For Hobbes, the sole basis of obligation was "covenant" or agreement: The process in which, by word or deed, a person undertakes to do or refrain from something. The reason why we have a general obligation not to kill our fellow citizens is that we have covenanted to obey a sovereign, and he (if he knows his business) will have prohibited killing. Though we have given the sovereign in this way a sort of blank check to generate obligations for us, there are limits on that, and self-defense is supposed to be one of those limits. Hobbes's answer to question 2, then, consists in showing that a person cannot be thought of as having entered into a covenant not to

19. See WESLEY NEWCOMB HOHFELD, *FUNDAMENTAL LEGAL CONCEPTIONS* 38-50 (Walter Wheeler Cook ed., 1923). The reason why I say "something like Hohfeldian privilege" will become clear below. See *infra* text accompanying notes 49-51.

defend himself or a covenant to obey any general rule to the extent that it includes a prohibition on self-defense.

Hobbes offers two arguments to this effect: (i) an argument based on the *unintelligibility* of any covenant purporting to renounce the liberty of self-defense; and (ii) an argument based on the *impossibility* of performing any such covenant.

A. *Unintelligibility of a Covenant Renouncing Self-Defense*

The first line of argument, the unintelligibility of a person's renunciation of his liberty of self-defense, is prominent in *Leviathan*. It starts from the premise that every covenant is a voluntary act, "and of the voluntary acts of every man, the object is some Good to himself."²⁰

[T]he motive, and end for which this renouncing and transferring of Right is introduced, is nothing else but the security of a mans person, in his life, and in the means of so preserving life, as not to be weary of it. And therefore if a man by words, or other signes, seem to despoyle himselfe of the End, for which those signes were intended; he is not to be understood as if he meant it, or that it was his will; but that he was ignorant of how such words and actions were to be interpreted.²¹

From this Hobbes argues:

And therefore there be some Rights, which no man can be understood by any words, or other signes to have abandoned, or transferred. As first a man cannot lay down the right of resisting them, that assault him by force, to take away his life; because he cannot be understood to ayme thereby, at any Good to himself.²²

The argument has an interesting provenance in political theory.²³ In the sixteenth and seventeenth centuries, a dispute emerged about the alienability of liberty (in relation to contractarian defenses of slavery and absolutism). Some (such as Gerson and Suarez) argued that liberty could be alienated as much as anything else one owned, if the price were right, and therefore we could concoct the possibility of a contractarian argument for slavery just by indicating its attractiveness as an option compared with other options faced by the subject.²⁴ Some (such as Francisco de Vitoria) opposed this in principle, insisting that our liberty (like many of our other rights) was held as a trust from God, and that we lacked the juridical power

20. HOBBS, *supra* note 2, at 93 (emphasis omitted).

21. *Id.* at 93-94.

22. *Id.* at 93.

23. This paragraph summarizes the complex and fascinating account in RICHARD TUCK, *NATURAL RIGHTS THEORIES: THEIR ORIGIN AND DEVELOPMENT* (1979).

24. *See id.* at 25-27, 54-55.

to sell or alienate what in fact belonged to the Almighty.²⁵ In between, an argument emerged, unfavorable to slavery and absolutism, based however not on juridical inalienability, but on what philosophers call "the principle of charity":

[N]o man can be supposed so void of common sense (unless an absolute Foole, and then he is not capable of making any Bargain) to yield himself so absolutely up to anothers disposal, or to renounce all hopes of safety or satisfaction in this life, or of further happiness in that to come. So that I conceive that even a Slave (Much more a servant hired upon certain Conditions) in the state of Nature, where he hath no civil power to whom to appeal for Justice, hath as much right as a Son or Child of the Family, to defend his life, or what belongs to him, against the unjust violence or Rage of his Master. . . . Since we have no notions of happiness but in life, nor is that farther than it is accompanied with some contentment of mind, no rational man can be supposed to consent to renounce all the pleasures and ends thereof, (and which onely make life desirable) much less the Right of living and preserving himself. . . .²⁶

Hobbes is clearly making use of a similar principle.²⁷ But the principle of charity is one whose application goes only to the issue of intelligibility, and the link between that and moral authority is not self-evident. Why do we insist that people may be bound only by covenants that are intelligible (to us)? Partly it is a matter of interpretation. If any provision is unclear, we can interpret it only by imputing a purpose to the covenant as a whole: An individual's obligations depend, for Hobbes, not so much on the exact words of his submission but "from the Intention . . . which . . . is to be understood by the End [thereof]."²⁸ But if the covenant is unintelligible in this sense, no intention or purpose can be imputed to it.

In addition, intelligibility is crucial to the reason why covenants create obligations. According to Hobbes, if someone gives up a right, his obligation not to exercise it is based, so to speak, on the duty of noncontradiction:

So that Injury, or Injustice, in the controversies of the world, is somewhat like to that, which in the disputation of Scholers is called Absurdity. For as it is there called an Absurdity, to contradict what one maintained in the Beginning: so in the world,

25. See *id.* at 46-47. Locke's account of the basis of nonaggression and of the prohibition on suicide in the *Second Treatise* is along these lines. See LOCKE, *TWO TREATISES*, *supra* note 2, at 271.

26. TUCK, *supra* note 23, at 155 (quoting JAMES TYRRELL, *PATRIARCHA NON MONARCHA* 4-5 (1681)).

27. In the final chapters of *Natural Rights*, Tuck traces the evolution of Hobbes's view in this direction and away from the more absolutist views of the other positions. See TUCK, *supra* note 23, at 119-42.

28. HOBBS, *supra* note 2, at 150.

it is called Injustice, and Injury, voluntarily to undo that, which from the beginning he had voluntarily done.²⁹

One has to be a bit careful when Hobbes talks about "Scholars," for it usually means he is making a joke.³⁰ But the point is evident enough. Breaking a contract is, in a sense, self-defeating. You made it for a reason, and so you are defeating that reason by backing out of it.³¹ If there is anything in this line of argument at all, it cannot be applied to covenants not to defend oneself. For on Hobbesian psychology, such contracts cannot have a point anyway, and so it can hardly be called self-defeating to violate them.

More substantively, "Covenants are in vain, and but Empty words,"³² so long as men do not keep to them. Their emptiness is a matter of concern not because nature abhors a vacuum, but because the point of making covenants is to ameliorate the awful situation of the war of all against all in the state of nature. It is not *entering into* covenants, as such, that offers amelioration, but people actually doing what they covenanted. The doing, then, completes the process without which the whole enterprise of covenanting would simply leave matters as bad as they were. Now, in Hobbesian psychology, individuals do not have an interest in ameliorating the awfulness of the state of nature, as such; they have that interest only to the extent that their own survival and well-being is improved. The point of a person's entering into the covenants to ameliorate the state of nature by constituting sovereignty and civil society is not peace and good order in themselves, but peace and good order so far as they contribute to his individual security. If social order seems to require my death, then it has no normative force whatever as far as I am concerned, for the only norms that Hobbesian man is capable of following are maxims of individual survival. A covenant not to defend myself would be a covenant to give up the only stake I have in this whole issue. We can hardly defend my obligation to perform such a promise, then, in terms of the desirability of amelioration, for such a concern will have no grip on me.

It seems, from all of this, that the Hobbesian account of self-defense is hostage to the peculiarities of his psychology. So long as people care ultimately about nothing other than their own preservation, the argument goes through. But if they care about other things, the argument collapses. (For example, someone may give up his right of self-defense in order to enhance the survival prospects of his child or of some work of art.) But we should distinguish the content from the structure of the psychology that is crucial here. According to Hobbes, there is a primeval motivation, *M*, such

29. *Id.* at 93.

30. *See, e.g., id.* at 14 (final paragraph).

31. For a critique of this reasoning, see GREGORY S. KAVKA, HOBBSIAN MORAL AND POLITICAL THEORY 315-37 (1986).

32. HOBBS, *supra* note 2, at 100.

that no human act is intelligible if it directly defeats *M*.³³ A covenant to defeat *M* is meaningless, and it cannot serve as the ground of an obligation to defeat *M*. No civil law, therefore, can require a person to defeat *M*, and any law must be implicitly qualified to the extent that its observance is likely to come into conflict with *M*. That is the structure of the link between Hobbesian psychology and justifications like self-defense. In addition, Hobbes held that for all human beings (except madmen) *M* was the preservation of one's own life. We may disagree. We may think this interpretation of *M* too narrow (so that we extend self-defense to include the defense of one's family), or we may think it too broad (because it derogates from the intelligibility of suicide or self-sacrifice). But in attacking Hobbes, we should ask ourselves whether our view of human motivation is so different from his (so complicated or variegated, for example) that nothing can be identified that plays the role *M* plays in his argument.

B. *Impossibility of a Covenant Renouncing Self-Defense*

In his earlier work, *De Cive*, Hobbes coupled the argument we have just considered with another one concerning the irresistibility of the urge to survive. There is, he argues, in every individual a certain threshold of fear such that "[w]hen a man is arriv'd to this degree of fear, we cannot expect but [that] he will provide for himself either by flight, or fight."³⁴ This is an expectation not only of reasonableness but above all of "natural necessity." It arises from the fact that on Hobbes's view we are mechanisms programmed to resist the immediate evils of death and bodily debilitation. So a contract not to defend oneself is void not only because it is unintelligible, but also because it involves a promise to perform the impossible: "Since . . . no man is tyed to impossibilities; they who are threatened either with death, (which is the greatest evil to nature) or wounds, or some other bodily hurts, and are not stout enough to bear them, are not obliged to endure them."³⁵ To be "tyed to impossibilities," Hobbes says, "is contrary to the very nature of compacts."³⁶ Since "ought" implies "can," it is not the case that anyone ought to endure a threat of death without resistance, and therefore no covenant could possibly generate such an obligation.

Moreover, such impossibility inevitably affects the view that the parties to such a covenant would take of its performance. It is part of the Hobbesian background that the character of human motivation is well known. So no one to whom such a promise of nonresistance was made could possibly have an expectation of its being fulfilled. From the point of

33. A stronger version would have it that no human action is intelligible except to the extent that it can be related to *M*, directly or indirectly. But a view this strong is not necessary here.

34. HOBBS, *supra* note 12, at 58.

35. *Id.* at 58-59.

36. *Id.* at 59.

view of the promisor, the alleged promise would be equally mysterious. Given the common knowledge just mentioned, there would be nothing to be gained by such a covenant to do the impossible. No one would plausibly give anything in return for such a hollow undertaking.³⁷

III

HOBBSIAN NECESSITY

I shall address the crucial question of the sort of exculpation that is involved in Hobbes's account (for example, justification or excuse) in Part IV. First, though, I want to look at the way in which this theory of self-defense might extend to cover cases of necessity.

One characteristic distinction between self-defense and necessity is that the former concerns a response to violent aggression whereas the latter concerns a response to circumstances that need not involve the aggression of a human being. Hobbes's account is often posed most dramatically in terms of a person's response to others who are trying to kill him. Yet it is

37. It is interesting that four hundred years later, John Rawls used these points as the basis of his contractarian argument against utilitarianism. The parties to the social contract, he says, must consider "the strains of commitment":

They cannot enter into agreements that may have consequences they cannot accept. They will avoid those they can adhere to only with great difficulty. Since the original agreement is final and made in perpetuity, there is no second chance. In view of the serious nature of the possible consequences, the question of burden of commitment is especially acute. A person is choosing once and for all the standards which are to govern his life prospects. Moreover, when we enter an agreement we must be able to honor it even should the worst possibilities prove to be the case. Otherwise we have not acted in good faith. Thus the parties must weigh with care whether they will be able to stick by their commitment in all circumstances.

JOHN RAWLS, *A THEORY OF JUSTICE* 176 (1971). (I have used this Rawlsian/Hobbesian argument as the basis of an argument that every social system claiming contractarian support must include basic welfare provision. See Jeremy Waldron, *John Rawls and the Social Minimum*, 3 J. APPLIED PHIL. 21 (1986); Jeremy Waldron, *Welfare and the Images of Charity*, 36 PHIL. Q. 463 (1986); see also JEREMY WALDRON, *LIBERAL RIGHTS* 225-70 (1993) (reprinting these two articles)).

Rawls believes that a social structure based on utilitarian principles may have drastically bad consequences for the lives of some people (trading off their suffering against the greater happiness of others), consequences that are bad enough to induce them in desperation to attack, undermine, or in other ways withdraw their support from that structure. Since people know this in advance and since no one knows who these unfortunates will be, no one can promise in good faith to support a social system so grounded. The strains of commitment imposed by such a covenant with utilitarianism are simply too great.

Rawls believes (I think wrongly, see Waldron, *John Rawls and the Social Minimum*, *supra*) that this is an argument not only against unqualified utilitarianism but in favor of his two principles of justice. Be that as it may, the Hobbesian tendency of the argument is clear. No social obligation to refrain from desperate remedies in desperate situations can be grounded in consent, for *either* such consent is unintelligible *or* it amounts to the promise of such impossible restraint that no one in their right mind would rely on it and no one in good faith should offer it. An argument along these lines will clearly provide a basis for self-defense since, given certain minimally plausible propositions of human psychology, no one could in good faith offer and no one in their right mind would accept (as a basis for their subsequent dealings) a person's promise not to defend himself *in extremis*. A law of homicide that did not provide for exculpation in such cases could not possibly claim consensualist or contractarian legitimacy.

easy to extend it to circumstantial necessity as well, along the following lines.

For Hobbes, the *summum malum* is death, particularly premature death; it does not matter for the purposes of his theory whether death is due to the violent aggression of others or to the absence of necessary resources. Shunning death *as such*, which might consist either in fighting off an attack or in foraging for food, is the irresistible impulse for the Hobbesian man. It is true that Hobbes concerned himself particularly with cases of capital punishment. But there are other ways in which the law or the sovereign may do something which *in effect* threatens our lives, and one of them is by commanding us to refrain from using other people's food without their consent even when that food is necessary (and physically available) for our survival.

Maybe there is a macho version of Hobbesian man who fears the indignity of homicide at another's hand much more than he fears dying quietly of starvation or hypothermia, and who can put up with those latter predicaments quite stoically. People resent, for example, the active aggression of others in a way that is not matched by their resentment of others' failure to rescue them.³⁸ Much has been made of this point in modern discussions of justice. F.A. Hayek argues that since the poverty which is the upshot of market processes for some people is not attributable to any person's action or decision, we cannot complain of it on grounds of justice in the way that we might complain of an intentional harm.³⁹ The macho version of Hobbesian man would help to make sense of this position. That individual could be expected to put up with poverty but could not be expected to refrain from defending himself against an attack, precisely because the latter, unlike the former, was an intentional assault upon his dignity. The example shows how curious human motivation has to be in order to make sense of the Hayekian position. Most people are not that macho. For most of us, there is no massive or qualitative difference between our impulse to avoid death in all its forms and our impulse to avoid being murdered, in particular. And if there is no such massive or qualitative difference, then the Hobbesian argument is going to entail that dire poverty is as difficult for people to put up with as an attack upon their lives.

We can put this point in the terms we used earlier (in Part II), as a possible controversy about the meaning of "*M*." For Hobbes, *M* is self-preservation. For others, however, even others who take a broadly Hobbesian view (and at the end of the Essay I shall argue that Professor

38. See P.F. STRAWSON, *FREEDOM AND RESENTMENT, AND OTHER ESSAYS* 5 (1974).

39. See 2 F.A. HAYEK, *LAW, LEGISLATION AND LIBERTY: THE MIRAGE OF SOCIAL JUSTICE* 64 (1976) ("It has of course to be admitted that the manner in which the benefits and burdens are apportioned by the market mechanism would in many instances have to be regarded as very unjust *if* it were the result of a deliberate allocation to particular people. But this is not the case.").

Kadish falls into this camp), *M* may be restricted to self-defense against aggression. The shape of the view is the same; but its content and implications are different.

IV

HOBBSIAN SELF-DEFENSE: JUSTIFICATION OR EXCUSE?

The Hobbesian argument about self-defense is a familiar one, and it poses a number of well-known difficulties for the rest of his political theory that I shall not go into here.⁴⁰ But I do want to consider its implications for the way in which *we* should think about self-defense (and necessity). There are two questions to address: (a) is Hobbesian self-defense a justification or an excuse?; and (b) if it is a justification, what sort of justification is it?

Let us begin by noting the implications of the Hobbesian approach for the hypothetical case with which we began Part I of this Essay. We imagined a terrorist and his hostage/shield both shooting at a police officer who was shooting at them. The question was: are any of these three exculpated from the charge of attempting to cause another's death? If (as we assumed) the terrorist had the option of surrender, and as long as he did not face immediate execution upon arrest, then it would seem he is as culpable on the Hobbesian account as he would be in a modern system of law. He cannot plead self-defense, since surrender would offer an alternative way, and indeed, unless he faces the imminent prospect of execution, an arguably safer way, of saving himself.

But for the police officer and the hostage, the Hobbesian implication would seem to be that both are exculpated. In the circumstances, neither can expect to survive without killing the other, and neither can be thought of as obligated to refrain from resisting in a case such as this. The hardest case (from our point of view) is that of the hostage. Part of our temptation to say that he is not exculpated is that he is not facing unlawful force. But, for Hobbes, as we have seen, the compelling nature of self-defense stems from the fact that my life is threatened. The juridical status of the threat is neither here nor there. Death is death, whether or not it is inflicted by a man in uniform.

Another part of our temptation to inculcate the hostage stems from the fact that the state of affairs in which he survives is associated with the state of affairs in which the terrorist escapes with him, and that is worse, from a social point of view, than the state of affairs in which the police officer survives and the other two are killed. Why is it worse? Because the terrorist is an aggressor and the hostage is (albeit involuntarily) implicated in

40. See, for example, HAMPTON, *HOBBS AND THE SOCIAL CONTRACT TRADITION*, *supra* note 18, at 132-88, for an excellent game-theoretic solution to some of the problems that have been thought to arise.

that aggression, and because the terrorist's survival poses a further threat to others in the future. What is more, the hostage knows that the situation in which he and the terrorist survive is worse (for society) than a situation in which both are killed.

But—and this is the crucial point—for *Hobbes*, *there is no such social perspective so far as the hostage is concerned*. The only thing at stake in the matter for the hostage is his own survival. He may want to give some thought to whether he will be one of the terrorist's later victims if they both get out of this immediate situation, though he will probably prefer to take his chances with that future threat than abandon the defense of his life here and now. If there is time for the question to flash through his mind, he may ask whether he is acting unjustly or breaking a social or political obligation by shooting at the police officer. But he will conclude that the only obligation he could be breaking would be a voluntary promise not to defend his life in circumstances where it was better for others that he sacrifice it. He will think that he could not have given such a promise, nor would such a promise be intelligible if given, nor would anyone expect him to carry it out if he had or rely on his performing it. He will conclude, then, that he is doing nothing wrong.

Since similar reasoning is available to the police officer (to explain why he is doing no wrong shooting to kill the innocent hostage who is threatening his life), it looks as though Hobbes's theory will entail the paradoxical result that two people trying to kill each other may both be exculpated on grounds of self-defense.

Notice that the paradox exists so long as the exculpation is thought of in terms of justification. How could two acts with exactly opposite tendencies both be *justified*? But there is no difficulty at all with the notion that two such acts might both be *excused* (or that one might be excused and the other justified). If a madman attacks me, believing crazily that I am trying to kill him, I may be justified in fighting back, even though he lacks the state of mind requisite for saying that he has done something wrong. If two madmen may set upon each other, each convinced that the other is a mortal threat, both may be excused for their attacks. The reason the paradox dissolves in such cases is that excuse is *obviously* agent-relative. To excuse an act is to say something about the agency-conditions of the person who performs it; it is not to say anything about the objective character of the act itself, its outcome, or its relation to any other agent. We are familiar with this in the idea that a sane accessory might be guilty of an offense of which the principal has been exculpated through some excuse.

Some of what Hobbes says about self-defense sounds like the language of excuse. When he characterizes one's response to fear in terms of

“natural necessity,” he seems to be talking in terms of irresistible impulse.⁴¹ Hobbes seems to be impugning the voluntariness of the behavior in question, by suggesting that there is no element of choice in the person’s performance of it. This seems to be connoted in the second of the arguments we considered for his claim that no one can be deemed to have undertaken not to defend himself, the argument based on the impossibility of performance.

But as I mentioned earlier, the appearance of excuse is diminished once we ask Hobbes not only why the liberty of self-defense is inalienable, but also why people have such a liberty in the first place. This brings us to question 1.⁴² When we raise question 1—Why does Hobbes suppose that people begin with a natural liberty of self-defense?—it becomes clear that the answer is going to take us directly to the foundation of the Hobbesian theory of motivation. All there is to Hobbesian motivation is self-preservation; that is the spring and fountain of all action, on his psychology: “[F]or every man is desirous of what is good for him, and shuns what is evill, but chiefly the chiefest of naturall evills, which is Death; and this he doth, by a certain impulsion of nature, no lesse then that whereby a Stone moves downward.”⁴³

To our ears, this passage can sound deterministic: It seems to present human nature in the light of a hard-wired survival machine. Now, there are reasons for and against the attribution of behavioral determinism to Hobbes. But it is worth noting that, in this context, no objectionable determinism need be inferred. Any theory of rational motivation will impute desires as well as beliefs to human agents, and it will present action as the deliberate choice of a course that best satisfies desires given beliefs. Few such theories have much to say about the provenance of desire; for the most part, desires are taken as given. The point is that the given-ness of desire does not in itself amount to determinism, and that is all Hobbes is presenting here. He is insisting that every human has an overriding natural motivation—to shun “the chiefest of naturall evills, which is Death.”⁴⁴ Nothing here implies that behavior thus motivated is mechanical or nonde-liberative. It is Hobbes’s view that we cannot choose but to be motivated by this desire, but it would be a severe libertarianism indeed that insisted that free will requires the agent to be able to choose what motivates him.

For our purposes, the important point is the pivotal character of this motivation in Hobbes’s theory. In terms of the discussion in Part III, motivation by fear of death is Hobbes’s “*M*”—it is the spring of all action for

41. See HOBBS, *supra* note 12, at 58 (“[F]or there is in every man a certain high degree of fear, through which he apprehends that evill which is done to him to be the greatest, and therefore by naturall necessity he shuns it all he can, and ‘tis suppos’d he can doe no otherwise.”).

42. That is, the first of the two questions set out *supra* pp. 718-718.

43. HOBBS, *supra* note 12, at 47.

44. *Id.*

Hobbes. His account of deliberation and of the will leaves no room for the view that an action so motivated is involuntary: "By this it is manifest, that not onely actions that have their beginning from Covetousnesse, Ambition, Lust, or other Appetites to the thing propounded; but also those that have their beginning from Aversion, or Feare of those consequences that follow the omission, are *voluntary actions*."⁴⁵

"Feare and Liberty are consistent," for Hobbes, and "[c]ovenants entred into by fear . . . are obligatory."⁴⁶ Indeed the latter has to be the case, for the only reason people have for entering the social contract is their terror of the continuation of the state of war.⁴⁷ Thus, rather than fear negating voluntariness, fear is seen as its basis in most cases of human choice and deliberation. Fear for one's life then, and the impossibility of not being motivated by it, can hardly operate as an excuse in Hobbesian jurisprudence.⁴⁸

These considerations are reinforced by what Hobbes says generally about the liberty of self-preservation in the state of nature. His basic position is that self-preservation is a natural right:

The Right of Nature, which Writers commonly call *Jus Naturale*, is the Liberty each man hath, to use his own power, as he will himselfe, for the preservation of his own Nature; that is to say, of his own Life; and consequently, of doing any thing, which in his own Judgement, and Reason, hee shall conceive to be the aptest means thereunto.⁴⁹

I have already mentioned that Hobbesian "rights" turn out to be more like privileges than like Hohfeldian claim rights. The right of nature is certainly not correlated with anyone else's duty to permit or refrain from interfering with the subject's self-preservation. It is a Hohfeldian privilege in the sense that the only thing it connotes about duty is the absence of any

45. HOBBS, *supra* note 2, at 45.

46. *Id.* at 97, 146.

47. See the discussion of "Common-wealth by Acquisition," *id.* at 138.

48. A more sophisticated theory might distinguish between two different ways in which fear motivates action. One is rational—as when a person weighs the implications of each course of action for his survival and decides after deliberation to choose one rather than the other. The other is impulsive—as when a person is so panicked or terrified by an imminent threat that his deliberative faculties are in effect overwhelmed. There is a vivid description of this sort of terror in HANNAH ARENDT, *THE ORIGINS OF TOTALITARIANISM* 437-68 (1973). Arendt writes, for example, of "the bestial, desperate terror which, when confronted by real, present horror, inexorably paralyses everything that is not mere reaction," and she suggests that "under conditions of total terror not even fear can any longer serve as an advisor of how to behave." *Id.* at 441, 467.

One cut on the distinction might be this: An action is motivated by terror rather than fear if its "choice" is impelled by circumstances despite the fact that deliberation motivated by survival might counsel an alternative option. Terror might then operate as an excuse—indicating that agential faculties were short-circuited or overwhelmed—in a way that fear might not. I have discussed this contrast in Jeremy Waldron, *Autonomy and Perfectionism in Raz's Morality of Freedom*, 62 S. CAL. L. REV. 1097, 1115 (1989).

49. HOBBS, *supra* note 2, at 91.

duty incumbent upon the subject not to defend himself. But, interestingly, it is not merely a Hohfeldian privilege. By calling self-defense a right, Hobbes wants to say more than that there is no opposing duty: He wants to commend self-defense as reasonable and maybe even prescribe it as requisite for rationality. It is, he says (in *De Cive*), "neither absurd, nor reprehensible; neither against the dictates of true reason for a man to use all his endeavours to preserve and defend his Body, and the Members thereof from death."⁵⁰ And in *Leviathan*, he elevates self-defense into natural law:

[I]t is a precept, or generall rule of Reason, That every man, ought to endeavour Peace, as farre as he has hope of obtaining it; and when he cannot obtain it, that he may seek, and use, all helps, and advantages of Warre. The first branch of which Rule, containeth the first, and Fundamentall Law of Nature; which is, to seek Peace, and follow it. The Second, the summe of the Right of Nature; which is, By all means we can, to defend our selves.⁵¹

The law of nature provides the ultimate terms of justification for Hobbes. If this is not a justification (as opposed to an excuse) in Hobbes's theory, nothing is.

Why do we find it difficult to regard this as a genuine theory of justification? The answer, I think, has to do with the agent-relative character of Hobbes's ethics. Hobbes is at pains to emphasize that, so far as the evaluation of actions is concerned, there is no *moral point of view* in the state of nature.⁵² There is nothing but the separate individual points of view of each of its inhabitants. That does not mean there is nothing to be said about justification; it is not a nihilist theory. But it means that everything that may be said about justification can be said only from the point of view of some one individual or other. There is no overarching or uniting point of view which could be the basis, for example, of arguing that one person's survival was a lesser or greater evil overall than that of another.

We must be careful not to misunderstand this doctrine. I said it is not nihilistic, and it is not relativistic either. I mean it is not relativistic in the sense in which moral relativists are sometimes taken to say (skeptically) that there is no moral truth. In this sense, the idea of relativism is quite different from the idea of agent-relativity. The passage that is often cited as a basis for attributing relativism to Hobbes is the following:

But whatsoever is the object of any mans Appetite or Desire; that is it, which he for his part calleth *Good*: And the object of his Hate, and Aversion, *Evill*; and of his Contempt, *Vile*, and *Inconsiderable*.

50. HOBBS, *supra* note 12, at 47.

51. HOBBS, *supra* note 2, at 91-92 (emphasis omitted).

52. See *id.* at 90 ("To this warre of every man against every man, this also is consequent; that nothing can be Unjust. The notions of Right and Wrong, Justice and Injustice have there no place. Where there is no common Power, there is no Law: where no Law, no Injustice.").

For these words of Good, Evill, and Contemptible, are ever used with relation to the person that useth them: There being nothing simply and absolutely so; nor any common Rule of Good and Evill, to be taken from the nature of the objects themselves.⁵³

But we should notice two points about this. First, Hobbes is talking here about the (attitudinal) meaning of the terms "good" and "evil," and it is clear from elsewhere in his argument that he holds strong views about the criteria of their application⁵⁴ (that is, about what *should be* the object of any man's appetite). Secondly, the passage indicates that "good" and "evil" should be regarded as indexical expressions (like "here" and "now"). It does not suggest that the meaning of the terms is up to the person. Hobbes is not saying that people disagree about what is justified and unjustified and that there is no way of resolving that disagreement. If someone (call him A) were to maintain that an action of his is justified if and only if it increases the amount of beauty in the universe, Hobbes would say categorically that he was wrong: The only thing that can justify an action by A is its conducing to A's survival. The theory, then, is a committed theory; it is not a theory of "anything goes." But its commitment is to an agent-relative proposition. It maintains categorically that, for each person X, the only thing that justifies an action by X is that action's conducing to X's survival. It rejects any other theory as mistaken.

The Hobbesian account rejects "anything goes" in another way as well. What is justified by the right of nature is what a man conceives "in his own Judgement, and Reason, . . . to be the aptest means"⁵⁵ to self-preservation. If a person acts in a way that his own reason counsels against in this regard, he can be criticized as foolish.⁵⁶ Of course the test remains what lawyers call "subjective":

Now whether the means which he is about to use, and the action he is performing, be necessary to the preservation of his Life, and Members, or not, he himself, by the right of nature, must be judg; for say another man, judg that it is contrary to right reason that I should judg of mine own perill: why now, because he judgeth of what concerns me, by the same reason, because we are equall by nature, will I judge also of things which doe belong to him; therefore it agrees with right reason (that is) it is the right of nature that I judge of his opinion, (i.e.) whether it conduce to my preservation, or not.⁵⁷

53. *Id.* at 39.

54. For the distinction between meaning and criteria, see R.M. HARE, *THE LANGUAGE OF MORALS* (1952).

55. HOBBS, *supra* note 2, at 91.

56. *See id.* at 101 (identifying as the "Foole" a person who acts contrary to his reasoned self-preservation).

57. HOBBS, *supra* note 12, at 47 (emphasis omitted).

But there is still a standard that people are to live up to. "Subjectivity" concerns only the issue of *who decides* whether the standard is satisfied; it doesn't affect the existence of the standard itself.

Clearly, then, there is a difference between a Hobbesian theory of justification and no theory of justification at all. What distinguishes Hobbes's theory is simply that it assigns different aims to different agents,⁵⁸ rather than assigning the same aim (for example, the welfare of mankind, or the social well-being) to all. Each agent is assigned the aim of his own self-preservation, and there is nothing more to justification than his actions serving that aim. It follows that this agent-relative theory of justification cannot choose, as between two people, whose survival is better, except in the trivial sense that *X's* survival is better for *X*, and *Y's* for *Y*. As we have found, this is really quite useful. What we wanted to say in the hypothetical case of the hostage/shield and the policeman is that both may be justified in attacking each other; we were *embarrassed* by our sense that we might have to choose.

Might someone not insist, nevertheless, that whatever the Hobbesian theory is, it is not a theory of *moral* justification, since it is indexical to agents in this way? Might someone insist that to be a *moral* theory, a theory must define a transcendent (or in our sense, agent-neutral) notion of justification? Well, we can say this if we like. The price is that a great many theories that have been regarded as rigorously moralistic will go out the window as well, in the face of such a requirement. There is no space to elaborate this here, but it is well established in ethics that most deontological theories of duty—particularly those that insist, for example, in a Kantian spirit, that one may not kill in order to save lives—are agent-relative.⁵⁹ These views are not egoistic (like Hobbes's view), but they are not agent-neutral either. And indeed there is a whiff of egoism in their suggestion that each agent must be overridingly concerned that evil should not come into their world *through him*.⁶⁰

V

JUSTIFICATION AND LEGITIMACY

There is an important feature of Hobbes's view that I have neglected until now. Partly as a consequence of its agent-relative structure, and partly for independent reasons, Hobbes's argument establishes only that the citizen does no wrong in resisting a violent attack; it does *not* establish that it

58. Please do not say, "Each agent is assigned the *same* end, namely self-preservation." That is a verbal confusion. Since the preservation of *X* is different from the preservation of *Y*, it follows that the aim assigned to *X* is different from the aim assigned to *Y*.

59. Robert Nozick's theory of rights is a modern example. See NOZICK, *supra* note 1, at 28. For a discussion, see NAGEL, *supra* note 1, at 175-85.

60. For a discussion, see Bernard Williams, *Utilitarianism and Moral Self-Indulgence*, in MORAL LUCK 40 (1981).

is wrong to violently attack *him*, or that he has a *right* not to be violently attacked. There is nothing in his theory comparable to the modern doctrine that an aggressor is not entitled to defend himself against the vigorous self-defense of his potential victim, for there is no way of establishing a right for one Hobbesian man that would have any impact on the issue of what it was right or wrong for any other man to do. Also, Hobbes's doctrine of self-defense does not establish that it is wrong for a sovereign to convict, condemn, or punish a person for defending himself. His argument establishes only that a person has no obligation to submit to punishment, for we cannot reasonably expect him to submit to it; the theory is not in a position to infer from this anything about the legitimacy of his being punished. It is a theory of exculpation, not indemnification.⁶¹

It is well-known that obligation and legitimacy can sometimes come apart in Hobbes's philosophy. For example, a group of people who have wrongfully rebelled against their sovereign may legitimately be put to death by him even though, for exactly the reasons we have been discussing, they have no obligation to go quietly to their execution or to refrain from conspiracy to escape.⁶² Legitimacy does not "ride in on the back of" obligation for Hobbes.⁶³

Why does the sovereign have a right to punish me for something that I had no obligation to refrain from, and in circumstances in which it is not reasonable for me to submit to the punishment? The answer is that in Hobbes's political theory, the actions of the sovereign are legitimated on grounds that are quite independent of the consent, covenants, and obligations of individual citizens. Their covenants concern their own obligations, and perhaps also the degree to which they must acknowledge the sovereign as their representative; the covenants limit their natural rights. But the natural right of the sovereign continues unabated—he does not give up any of his rights in the process of social contract—and that is the basis on which he may do anything to the citizen that he conceives necessary for his own (that is, the sovereign's) self-preservation. As Hobbes puts it:

[B]efore the Institution of Common-wealth, every man had a right to every thing, and to do whatsoever he thought necessary to his own preservation; subduing, hurting or killing any man in order thereunto. And this is the foundation of that right of Punishing,

61. This might seem bewildering: What could "exculpation" mean if not that punishment is inappropriate? But I am trying to use terms precisely. "Exculpation" means being free from *blame* (or if it does not, imagine a synonymous term that does), and as we know (especially in a Hobbesian context), there are sometimes reasons for punishing people apart from their being blameworthy for some offense.

62. See HOBBS, *supra* note 2, at 152. For a discussion of the gap between obligation and legitimacy, see Jeremy Waldron, *Theoretical Foundations of Liberalism*, 37 PHIL. Q. 127, 136 (1987), reprinted in LIBERAL RIGHTS, *supra* note 37, at 35.

63. The phrase is from Ronald Dworkin's discussion of legitimacy and obligation in LAW'S EMPIRE 191 (1986).

which is exercised in every Common-wealth. For the Subjects did not give the Sovereign that right; but onely in laying down theirs, strengthened him to use his own, as he should think fit . . . so that it was not given, but *left* to him, and to him onely . . . as entire, as in the condition of meer Nature, and of warre of every one against his neighbour.⁶⁴

It may be silly or counter-productive for the sovereign to prohibit self-defense, and in one passage, Hobbes indicates as an implication of his argument, that no threat of penalty could possibly dissuade the victim of an attack from fighting for his life: “[M]an by nature chooseth the lesser evill, which is danger of death in resisting; rather than the greater, which is certain and present death in not resisting.”⁶⁵ But the sovereign can have ulterior motives for the prohibition, and there is no basis in Hobbes’s theory for criticizing that.

I believe this is a feature of Hobbes’s approach to political philosophy that can be bracketed out of the rest of his discussion of self-defense. The crucial move would be to insist that the powers of the sovereign body are conferred rather than natural.⁶⁶ One would then have a basis for a contractarian theory of legitimacy, not just a contractarian theory of obligation. It would follow that what it was permissible for the sovereign to do would be subject, as much as the doctrine of obligation, to considerations about what it was reasonable for subjects to put up with. (As we shall see, this is one of the significant differences between Hobbes’s theory and Sanford Kadish’s Hobbesian theory.)

VI

LOCKE ON SELF-PRESERVATION

We have seen that Hobbes’s theory of self-defense is a theory of agent-relative justification. There is an interesting contrast, in this respect, with the account given by John Locke. Locke’s theory of self-defense is also a theory of justification (as opposed to excuse). But it purports, at least in its basis, to be agent-neutral rather than agent-relative in character. According to Locke, when someone declares

by Word or Action, not a passionate and hasty, but a sedate settled Design, upon another Mans Life, [he] puts him[self] in a State of War with him against whom he has declared such an Intention, and so has exposed his Life to the others Power to be taken away by him, or any one that joyns with him in his Defence, and espouses his Quarrel: it being reasonable and just I should have a Right to destroy that which threatens me with Destruction. For by the

64. HOBBS, *supra* note 2, at 214 (emphasis added).

65. *Id.* at 98.

66. One would have to say this anyway for the cases—allowed by Hobbes, *see, e.g., id.* at 129–38—in which the sovereign is not a natural person.

Fundamental Law of Nature, Man being to be preserved, as much as possible, when all cannot be preserv'd, the safety of the Innocent is to be preferred⁶⁷

The last sentence by itself is enough to distinguish the Lockean theory from the Hobbesian. It is impersonal in tone—"the safety of the Innocent is *to be preferred*"—and the law of nature counsels the preservation of mankind as such, rather than counseling every man to prefer his own survival. Talking about the safety of the innocent being (impersonally) *preferred* to the safety of the aggressor strikes a chord congenial to our modern theories of the "lesser evil."

Why does innocence have this superiority? At first sight, it looks as if we are dealing here with what Judith Jarvis Thomson calls a theory of "forfeit": "Aggressor, by virtue of his attack on Victim, has forfeited his right to not be killed, and therefore his right to life."⁶⁸ Certainly Locke presents his point in terms of a very stark contrast between guilt and innocence. Innocence is associated with humanity, aggressive guilt with bestiality:

one may destroy a Man who makes War upon him, or has discovered an Enmity to his being, for the same Reason, that he may kill a Wolf or a Lyon; because such Men are not under the ties of the Common Law of Reason, have no other Rule, but that of Force and Violence, and so may be treated as Beasts of Prey⁶⁹

That theme is also developed in Locke's account of punishment. Though he emphasizes that punishment is limited to "so much as may serve for Reparation and Restraint" or "as may make [the offender] repent the doing of it, and thereby deter him, and by his Example others, from doing the like mischief,"⁷⁰ he also stresses that the basis of punishment is a man's right to destroy or restrain noxious animals. By deviating from the law of nature, "a Man so far becomes degenerate, and declares himself to quit the Principles of Human Nature, and to be a noxious Creature."⁷¹ A murderer, for example, "having renounced Reason, the common Rule and Measure, God hath given to Mankind, hath by the unjust Violence and Slaughter he hath committed upon one, declared War against all Mankind, and therefore

67. LOCKE, TWO TREATISES, *supra* note 2, at 278-79 (emphasis omitted).

68. Judith Jarvis Thomson, *Self-Defense and Rights*, in RIGHTS, RESTITUTION, AND RISK: ESSAYS IN MORAL THEORY 34 (1986).

69. LOCKE, TWO TREATISES, *supra* note 2, at 279 (emphasis omitted). The bestial imagery is pervasive in the *Two Treatises* and is savagely directed against Locke's regal opponents, of whom he says, in the epigraph to the work: *Placari nequeant, nisi hauriendum sanguinem laniandaque viscera nostra praeberimus*. (Trans: "They will not be satisfied unless we yield to them our blood to drink and our entrails to be torn out."). *Id.* at 136.

70. *Id.* at 272 (emphasis omitted).

71. *Id.* at 273.

may be destroyed as a Lyon or a Tyger, one of those wild Savage Beasts, with whom Men can have no Society nor Security."⁷²

But Locke's theory is not entirely one of forfeiture. Aggressors are not put *exactly on par* with wild beasts in Locke's moral ontology. Though we have a right to destroy them, we are not authorized to *use* them as we are "the inferior ranks of creatures."⁷³ As we have just seen, Locke imposes limits on what one can do to an aggressor in the way of punishment, and those limits make no sense if one has forfeited the rights that would otherwise act as a constraint. Moreover, a pure doctrine of forfeit would be incompatible with what Locke later says about the prerogative power of pardon: "For the end of Government being the preservation of all, as much as may be, even the guilty are to be spared, where it can prove no prejudice to the innocent."⁷⁴

This latter passage seems to indicate that the real reason for preferring the safety of the innocent to that of the aggressor is the continued threat that aggressors usually pose. In most circumstances, the survival of the guilty amounts to a threat not only to the particular innocent person whom he confronts, but also to other innocent people in the future. The emphasis Locke places on the bestiality of aggression is important not only as a theory of status-degradation or forfeiture, but also as an indication of his (perhaps false) belief that "Once an aggressor—always an aggressor." The aggressor "has quit the Principles of Human Nature" and so he becomes dangerous not only to the one he attacks but to mankind in general. It is better, then, from the point of view of the preservation of mankind, that this "noxious Creature" be destroyed, as one might destroy "a Lyon or a Tyger, one of those wild Savage Beasts, with whom men can have no Society nor Security."⁷⁵

It follows that the victim of an attack, who fights back in self defense, is acting on account of all who might be threatened in the future, rather than simply for himself. This is consonant with the distinction Locke makes between the natural right of punishment and the natural right of reparation. The latter resides solely in the injured party, whereas the former might be exercised by the victim or anyone.⁷⁶ True, punishment is not self-defense even for Locke, but the two are sufficiently close in purpose, on his account, to sustain the point.

72. *Id.* at 274 (emphasis omitted).

73. *Id.* at 271 ("And being furnished with like Faculties, sharing all in one Community of Nature, there cannot be supposed any such Subordination among us, that may Authorize us to destroy one another, as if we were made for one anothers uses, as the inferior ranks of Creatures are for ours.") (emphasis omitted).

74. *Id.* at 375 (emphasis omitted).

75. *Id.* at 274 (emphasis omitted).

76. *See id.* at 273-74.

In Hobbes's theory, we saw that action in the state of nature was primevally motivated by the desire to stay alive. That is the basis on which people act, and their conduct motivated in that way must always be regarded as reasonable. Is there an equivalent for Locke to this fundamental principle of Hobbesian natural right? There is, and as one would expect, it is considerably less egoistic:

For in the State of Nature, to omit the liberty he has of innocent Delights, a Man has . . . [the power] . . . to do whatsoever he thinks fit for the preservation of himself and others within the permission of the Law of Nature: by which Law common to them all, he and all the rest of Mankind are one Community, make up one Society distinct from all other Creatures. And were it not for the corruption, and vitiousness of degenerate Men, there would be no need of any other; no necessity that Men should separate from this great and natural Community, and by positive agreements combine into smaller and divided associations.⁷⁷

Notice that this theory, as much as Hobbes's, provides a subjective test of good behavior: The power is to do "*whatsoever he thinks fit*" for the preservation of mankind. But unlike Hobbes's theory, it is agent-neutral; that is, everyone is assigned exactly the same principle of action. Just as Hobbes's justification of self-defense is derived fairly directly from his agent-relative principle of reasonable action, so Locke's justification is derived fairly directly from his agent-neutral account. The preservation of mankind is the overarching principle of the law of nature, and the justification of self-defense is simply a special application of that.

The agent-neutral character of Locke's theory is bolstered, of course, by a theological dimension largely lacking in Hobbes's account. That as many humans as possible are to be preserved is understood as a matter of respect for God's creation. He created us for a purpose, evidently intending us to "live and abide for some time upon the Face of the Earth, and not . . . [to] . . . perish again, presently after a few moments continuance," making us "to last during his, not one anothers Pleasure."⁷⁸ Allegiance to this purpose is the basis of our natural duty not to harm one another and our natural duty to see to one another's survival as best we can.

In a recent book, George Fletcher has suggested that Locke's theory of self-defense "takes the vindication of autonomy as its fundamental imperative," and that this "individualist" theory of self-defense stands in marked contrast to the social theory of the lesser evil.⁷⁹ Certainly, Locke says that I may resist anyone who attacks me as a threat to my liberty. Yet

77. *Id.* at 352 (emphasis omitted). This passage, like many in Locke, *see, e.g., id.* at 318-19, is quite incompatible with the atomism usually attributed to him.

78. *Id.* at 205, 271.

79. GEORGE P. FLETCHER, A CRIME OF SELF-DEFENSE: BERNARD GOETZ AND THE LAW ON TRIAL 32-33 (1988).

he insists in the same breath that the reason I may resist a threat to my liberty is that it amounts to a potential threat to my life. It is quite impossible to read the passages in question as making the vindication of autonomy fundamental.⁸⁰ Elsewhere, Fletcher suggests that, according to Locke, "aggression breaches an implicit contract among autonomous agents, according to which each person or country is bound to respect the living space of others."⁸¹ That is an interesting idea, but it is not Locke's. The wrongness of aggression has nothing to do with Locke's contractarianism. It is prior to that, and rests on the theological premise that we are God's property, "made to last during his, not one another's Pleasure."⁸²

VII

RATIONALITY AND THE IMPULSE TO SELF-DEFENSE

The Lockean case for self-defense is supposed to be rationalistic. The idea is that it can be figured out as a matter of natural law by anyone who reflects on the human condition, the ground of his being, and the nature of aggressive dispositions. But Locke is also realistic enough to know that people's impulse to defend themselves has a nonrational grounding as well. In this part of the Essay, I examine various possible understandings of the place of such an impulse in Lockean theory.

In Section 11 of the *Second Treatise*, John Locke adds the following comment to his discussion of the natural right to destroy aggressors as noxious animals:

And Cain was so fully convinced, that every one had a Right to destroy such a Criminal, that after the Murder of his Brother, he cries out, Everyone that findeth me, shall slay me; so plain was it writ in the Hearts of all Mankind.⁸³

This passage is significant in the Locke literature mainly as a basis for attributing to him philosophical inconsistency, and as evidence for the claim that he did not wear his philosophical cap when he wrote *Two*

80. See LOCKE, *TWO TREATISES*, *supra* note 2, at 279-80 (emphasis omitted):

To be free from such force [sc. force attacking my freedom] is the only security of my Preservation: and reason bids me look on him, as an Enemy to my Preservation, who would take away that Freedom, which is the Fence to it; . . . using force, where he has no Right, to get me into his Power, let his pretence be what it will, I have no reason to suppose, that he, who would take away my Liberty, would not when he had me in his Power, take away every thing else.

81. FLETCHER, *RETHINKING CRIMINAL LAW*, *supra* note 8, at 861.

82. LOCKE, *TWO TREATISES*, *supra* note 2, at 271. Professor Fletcher fails to see that Locke's contractarianism has the feature that many of a person's rights are defined as natural rights, that is, as background to the social contract. This is in contrast with a Rawlsian theory in which all rights are determined by the contract (except of course the right to be treated as an equal in the determination of rights). See RAWLS, *supra* note 37, at 118; see also RONALD DWORKIN, *TAKING RIGHTS SERIOUSLY* 168-83 (1977) for a helpful discussion of the role of contract in Rawls's rights theory.

83. LOCKE, *TWO TREATISES*, *supra* note 2, at 274 (emphasis omitted).

Treatises of Government.⁸⁴ The passage seems to evince what Peter Laslett has called "the naive form"⁸⁵ of the belief in innate ideas, a belief that Locke excoriated in all its forms in Book One of the *Essay Concerning Human Understanding*,⁸⁶ published at roughly the same time. It may be better, however, to regard the passage as a simple slip, born of zealous rhetoric, rather than as an indication that Locke abandoned his philosophical views when writing about politics.⁸⁷ (Locke, like many philosophers since, loosens his intellectual standards when it comes to arguing in favor of capital punishment.) Certainly he makes nothing, in this context, of these innatist connotations. In all his discussion of self-defense and the law of nature generally, preceding and following this passage, he stresses the rationalist aspect.

The passage about Cain suggests that aggressors intuit that their potential victims have a right to kill them. That is less plausible than the converse claim that potential victims have a good sense, prior to rational calculation, that they have a right to defend themselves. Interestingly, Locke does not defend the latter claim in the *Second Treatise*. He insists that people know about the right of self-defense independently of being taught about it by lawyers, but that is because the reasoning that leads to it is pretty obvious and straightforward, not because it is innate or intuitive.⁸⁸ Such parts of the law of nature are

as intelligible and plain to a rational Creature, and a Studier of that Law [sc. the law of nature], as the positive Laws of Commonwealths, nay possibly plainer; As much as Reason is easier to be understood, than the Phansies and intricate Contrivances of Men, following contrary and hidden interests put into Words; For so truly are a great part of the Municipal Laws of Countries, which are only so far right, as they are founded on the Law of Nature, by which they are to be regulated and interpreted.⁸⁹

84. *Id.*

85. *Id.* at 274 (Laslett's footnote); see also *id.* at 79 (Laslett's introduction). The theory of innate moral ideas is attacked in LOCKE, HUMAN UNDERSTANDING, *supra* note 2, at Book I, ch. 3. For Laslett's claim that Locke the political theorist is different from Locke the philosopher, see LOCKE, TWO TREATISES, *supra* note 2, at 92-106 (section IV.2 of his editorial introduction). This is part of the general program of Laslett (and Quentin Skinner, among others) to cast doubt on the enduring philosophical significance of historically located and politically motivated works of political theory.

86. LOCKE, HUMAN UNDERSTANDING, *supra* note 2.

87. See IAN HARRIS, THE MIND OF JOHN LOCKE 31 (1994), for the welcome suggestion that the passage is not innatist at all, and that something can be writ in the hearts of all mankind by their own reasoning.

88. When he links the right of self-defense to that of revolution at the end of the work, Locke asks: "Are the People to be blamed, if they have the sence of rational Creatures, and can think of things no otherwise than as they find and feel them?" LOCKE, TWO TREATISES, *supra* note 2, at 418; see also *id.* at 415.

89. *Id.* at 275 (emphasis omitted).

It is when potential victims rely on their passions and instincts that things are likely to get out of hand. "Self-love will make Men partial to themselves and their Friends. And on the other side, . . . Ill Nature, Passion and Revenge will carry them too far in punishing others."⁹⁰

In the *First Treatise*, however, Locke does discuss an instinct of self-preservation (though he does not relate it explicitly to self-defense):

God having made Man, and planted in him, as in all other Animals, a strong desire of Self-preservation, and furnished the World with things fit for Food and Rayment and other Necessaries of Life, Subservient to his design, that Man should live and abide for some time upon the Face of the Earth, and not that so curious and wonderful a piece of Workmanship by its own Negligence, or want of Necessaries, should perish again presently after a few moments continuance: God, I say, having made Man and the World thus, spoke to him, (that is) directed him by his Senses and Reason, as he did the inferior animals by their Sense, and Instinct, which he had placed in them to that purpose, to the use of those things, which were serviceable for his Subsistence, and given him as means of his Preservation. And therefore I doubt not, but before these words were pronounced, I Gen. 28, 29. (if they must be understood Literally to have been spoken) and without any such Verbal Donation, man had a right to a use of the Creatures, by the Will and Grant of God. For the desire, strong desire of Preserving his Life and Being having been Planted in him, as a Principle of Action by God himself, Reason, which was the Voice of God in him, could not but teach him and assure him, that pursuing that natural Inclination he had to preserve his Being, he followed the Will of his Maker, and therefore had a right to make use of those Creatures, which by his Reason or Senses he could discover would be serviceable thereunto. And thus Man's Property in the Creatures, was founded upon the right he had, to make use of those things, that were necessary or useful to his Being.⁹¹

Humans, we are told, have a "natural inclination" to preserve their existence and to make use of the means that are evidently available for that. They can feel their hunger,⁹² see the food that exists on the earth, and

90. *Id.*

91. *Id.* at 204-05 (emphasis omitted). I have quoted the passage at length, because of its considerable philosophical interest. It weaves together considerations of reason and inclination in a powerful and subtle argument.

92. In *Human Understanding*, *supra* note 2, Book IV, ch. 10, § 2, Locke remarks: "If anyone pretends to be so sceptical as to deny his own existence (for really to doubt of it is manifestly impossible), let him for me enjoy his beloved happiness of being nothing, until hunger or some other pain convince him of the contrary." Man's "profane stomach," as Marx put it, convinces him that he is not merely a rational existent, but a being with needs that will not survive unless he is fed (and provided other necessities). See Karl Marx, *The Holy Family*, in KARL MARX: SELECTED WRITINGS 148 (D. McLellan ed., 1977). On the basis of our experience, we learn that the clamouring of a human body ends in death and decay if it is not attended to; and from the same source we learn that pangs of

experience the urge to take advantage of its availability. Locke uses "natural inclination" arguments a number of times in his discussion of natural law. He uses it, for example, in his rationalization of inheritance: "God planted in Men a strong desire also of propagating their Kind, and continuing themselves in their Posterity."⁹³

This mode of argument raises interesting issues of moral epistemology. On the one hand, it *sounds* as though Locke is committing himself to some doctrine of innate ideas, so far as natural law is concerned: We are born with the idea that self-preservation and propagation of the species are right. Alternatively, it can sound as though he is committing himself to some version of ethical naturalism: The rightness of some action is simply inferred from the existence of a natural impulse to perform it. Either way, the equation of natural instinct with moral rectitude poses difficulties for Locke's meta-ethical theory.

The difficulty with innate ideas is obvious enough. Locke spends a great deal of energy at the beginning of the *Essay Concerning Human Understanding* denying that there are any innate ideas, about morals or anything else: "[M]oral principles require reasoning and discourse, and some exercise of the mind, to discover the certainty of their truth. They lie not open as natural characters engraven on the mind."⁹⁴ We have already seen that there is one passage in the *Treatises* which appears flatly to contradict this considered epistemological position: the passage about Cain.⁹⁵ In his edition of the work, Peter Laslett puts the present passage on par with this slip, arguing that both are equally inconsistent with Locke's rejection of innatist epistemology.⁹⁶ But Laslett has not read the passage carefully enough.

What is striking about it is that Locke does *not* infer rightness directly from natural inclination. He says that "Reason . . . could not but teach him and assure him, that [in] pursuing that natural Inclination . . . he followed the Will of his Maker."⁹⁷ We find by reflection that we have this inclination, and we reason about ourselves and the world to find out what to make of it. The implicit assumption is that natural inclinations have to be scrutinized and reflected upon by reason before they yield any reliable moral

hunger can be assuaged by the use of food, the pain of cold by the use of material shelter and clothing, and so on. Through our senses, we learn that external resources are apt for human use, and that we need to use them in order to remain existent.

93. LOCKE, TWO TREATISES, *supra* note 2, at 206; *see also* JEREMY WALDRON, THE RIGHT TO PRIVATE PROPERTY 241-51 (1988); Jeremy Waldron, *Locke's Account of Inheritance and Bequest*, 19 J. HIST. PHIL. 39 (1981).

94. LOCKE, HUMAN UNDERSTANDING, *supra* note 2, at 66.

95. *See* LOCKE, TWO TREATISES, *supra* note 2, at 274; *see also supra* text accompanying notes 83-85.

96. *See id.* at 205 (Laslett's footnote); *id.* at 79-92 (section IV.2 of Laslett's editorial introduction).

97. *Id.* at 205.

guidance. Our reason tells us that this natural inclination is oriented to what can rationally be known about God's will in placing us here—"that Man should live and abide for some time upon the Face of the Earth," and what can be known experientially about how that will might be best served (by eating, for example).

It is not merely in the *Essay* that Locke denies we can infer a principle of right from a natural inclination. He denies that in the *Treatises* as well. In several passages he insists that humans have all sorts of inclinations, some of them nice, some of them nasty:

Nor can it be otherwise in a Creature, whose thoughts are more than the Sands, and wider than the Ocean, where fancy and passion must needs run him into strange courses, if reason, which is his only Star and compass, be not that he steers by.⁹⁸

As beings endowed with reason, we have to sort through our impulses and relate to some rational understanding of our being and some experiential knowledge of our nature. Then and only then are we in a position to draw normative conclusions from the fact of their existence.

It is worth noting, moreover, that this is almost exactly the point Locke makes in the discussion of innate ideas in the *Essay*. His discussion there does not deny the existence of natural inclinations:

I deny not that there are natural tendencies imprinted on the minds of men, and that from the very first instances of sense and perception, there are some things that are grateful and others that are unwelcome to them, some things that they incline to, and others that they fly: but this makes nothing for innate characters on the mind, which are to be principles of knowledge, regulating our practice. . . . Principles of actions indeed there are lodged in men's appetites, but these are so far from being innate moral principles that, if they were left to their full swing, they would carry men to the over-turning of all morality.⁹⁹

What we need, before we think it right to follow such an innate appetite, is some knowledge that it guides us as we ought to be guided. That knowledge, Locke insists, can only come from reason.

Although, as I have said, Locke does not himself apply this argument to self-defense, it is easy to see how it could be applied. Anyone who is attacked will impulsively defend himself, since there is this motivation, this "strong desire of Preserving his Life and Being having been Planted in him, as a Principle of Action by God."¹⁰⁰ And a moment's reflection, either at the time, or earlier or later in safety, will convince him that the impulse is rationally defensible.

98. *Id.* at 182.

99. LOCKE, HUMAN UNDERSTANDING, *supra* note 2, at 67.

100. LOCKE, TWO TREATISES, *supra* note 2, at 205.

Clearly the impulse to defend oneself plays a somewhat different role in the understanding of the right of self-defense in Locke from the role it plays in Hobbes. For Hobbes, the impulse is part of the justification: Since we are hard-wired to defend ourselves, it cannot be said that defending ourselves is wrong. But for Locke, the hard-wiring is, at most, an indication (and a defeasible one) that self-defense is justified. If our maker has wired us up in this way, he has presumably done so for a purpose, and rational reflection suggests that that purpose is self-defense. The rational justification matches all or most of the cases in which the motivation is felt. However, there is nothing inevitable about that for Locke. We are hard-wired for intense sexual desire also, but acting on that impulse is not justified in every occasion (or even most occasions) in which it is triggered. The rational justification for the indulgence of sexual desire indicates a moral limitation on the type of occasion in which it should be indulged.

So we have Hobbes, for whom the impulse is justification; and Locke, for whom the impulse is reliably correlated with the justification. A third possibility, which neither adopts, is that the impulse is a state of knowledge about the justification. That would be (a version of) the innate ideas approach wrongly imputed to Locke by Laslett (the true converse of the aggressor's knowledge slip in the passage about Cain). But Locke never takes that approach.¹⁰¹

VIII

SELF-DEFENSE AND NECESSITY

I mentioned earlier that with Hobbes there is a natural connection between the application of his theory to self-defense and its application in cases of natural necessity. There is a similar connection in Locke's philosophy. Locke is well known for the view that private property rights are based on individual labor. Whatever a man

removes out of the State that Nature hath provided, and left it in, he hath mixed his Labour with, and joyned to it something that is his own, and thereby makes it his Property. It being by him removed from the common state Nature placed it in, hath by this labour something annexed to it, that excludes the common right of other Men.¹⁰²

101. Maybe his theory implies it. Someone may argue for this as follows. If experiencing the inclination to defend oneself is a state everyone gets into in (almost) all and (almost) only the cases where such defense is justified, then it stands in a relation to the fact that defense is justified somewhat analogous to the relation between one's experiencing an object as red and its actually being red. Thus, that a situation is one in which a person is entitled to defend himself may be thought to be one of its secondary qualities.

102. LOCKE, *TWO TREATISES*, *supra* note 2, at 288 (emphasis omitted). For a critique, see WALDRON, *THE RIGHT TO PRIVATE PROPERTY*, *supra* note 93, at 171-207.

Locke emphasizes the implication of this doctrine, that "God gave the World . . . to the use of the Industrious and Rational (and Labour was to be his Title to it;) not to the Fancy or Covetousness of the Quarrelsome and Contentious."¹⁰³ But like most early modern theorists of political economy,¹⁰⁴ he qualified these property rights with an unequivocal provision for necessity:

God the Lord and Father of all, has given no one of his Children such a Property, in his peculiar Portion of the things of this World, but that he has given his needy Brother a Right to the Surplusage of his Goods; so that it cannot justly be denied him, when his pressing Wants call for it. . . . As Justice gives every Man a Title to the Product of his honest Industry, and the fair Acquisitions of his Ancestors descended to him; so Charity gives every Man a Title to so much out of another's Plenty, as will keep him from extrem want, where he has no means to subsist otherwise. . . .¹⁰⁵

This qualification is utterly in keeping with the basis of Locke's approach to the preservation of human life. The premise, as always, is that God intends the resources of the world to be used for human survival. Inspection of that idea (which Locke conducted at extraordinary length in the *First Treatise*) establishes that human survival includes the survival of all humans, or at least as many as possible. Introduce now the idea of a person, a "Man of Estate," in possession of a sufficient quantity of resources to keep not only himself alive, but others as well, and the idea of a "Needy Beggar," a poor man, with no resources of his own at all and no opportunity to acquire any from nature, the idea of a creature of God who will perish unless he makes use of resources owned by somebody else. If we put these two ideas together, we get the idea of a creature of God whose only possibility of survival is his being sustained by goods which are the property of another but which are not necessary to the latter's survival. From this we can see that the idea of the poor man's use of the rich man's surplus property is contained in the fundamental idea of God's intention that the resources of the world be used in a way that allows as many people as possible to survive.

IX

LOCKE AND THE TERRORIST EXAMPLE

What does Locke's theory imply for the hypothetical case stated at the beginning of Part I of this Essay? As with the other approaches, his account will have no difficulty with the idea that people are justified in

103. LOCKE, *TWO TREATISES*, *supra* note 2, at 291 (emphasis omitted).

104. See *WEALTH AND VIRTUE* 1-25 (Istvan Hont & Michael Ignatieff eds., 1983) (editors' introduction).

105. LOCKE, *TWO TREATISES*, *supra* note 2, at 170 (emphasis omitted).

shooting the terrorist and that the latter is not in justified shooting back at them. The terrorist "may be destroyed as . . . one of those wild Savage Beasts, with whom Men can have no Society nor Security."¹⁰⁶ But clearly the hostage cannot be shot at on this ground; he has not "renounced Reason"¹⁰⁷ nor "quit the Principles of Human Nature."¹⁰⁸ The ground for shooting at him is simple necessity: An innocent must perish so that others may survive.¹⁰⁹ Moreover, in relation both to the terrorist and the hostage, the police officer will know that his instinct of self-preservation is a reliable guide to what is right for him to do.

The difficulty as always is the hostage's point of view. If it really is the case that the terrorist's escape poses a threat to many more lives than the two that would be destroyed in eliminating him, and if the hostage knows this, it would seem that, on Lockean premises, he has no right to defend himself. For by defending himself, he is making it much more likely that many more people will perish. (He would be like a property holder resisting the destruction of his own home when that was necessary to stop a fire spreading.) Of course, the hostage's *impulse* will be to preserve himself. But that just reinforces the point made earlier: We cannot automatically rely on our impulses, since it is impossible for us to be hard-wired to make the fine discriminations that reason may sometimes demand. If the impulse is sufficiently compelling, maybe we can say that the hostage has an excuse for firing back at the police officer. But there is nothing so far in our account of Locke's theory to provide the basis for any claim of *justification*.

The situation is complicated, however, by two further points. First, Locke never actually discusses the sort of situation we are talking about. He never suggests that the law of nature may actually require self-sacrifice.¹¹⁰ It does not follow that our account is speculative: What we are doing is applying the deep logic of his position to this particular case.

Secondly, when he first introduces the notion of the preservation of mankind as the fundamental principle of natural law, Locke seems to formulate it in a way that gives priority for each person to his own survival: "Every one as he is bound to preserve himself, and not to quit his

106. *Id.* at 274.

107. *Id.* at 273.

108. *Id.* at 274.

109. See *id.* at 375 ("For since many accidents may happen, wherein a strict and rigid observation of the Laws may do harm; (as not to pull down an innocent Man's House to stop the Fire, when the next to it is burning).").

110. Cf. Michael Walzer, *The Obligation to Die for the State*, in OBLIGATIONS: ESSAYS ON DISOBEDIENCE, WAR, AND CITIZENSHIP 88-89 (1970) ("John Locke, because of the peculiarly collectivist formulation he gives to the fundamental law of nature, is able to reassert an obligation to die . . ."). But the passage that Walzer cites (*Two Treatises*, II, § 139)) concerns the legitimacy of martial discipline rather than the obligations of soldiers. See LOCKE, *TWO TREATISES*, *supra* note 2, at 362.

Station wilfully; so by the like reason *when his own Preservation comes not in competition*, ought he, as much as he can, to preserve the rest of Mankind."¹¹¹

The passage I have emphasized is ambiguous. It suggests but does not insist that a person has no obligation to preserve mankind when his own preservation is in question;¹¹² but all it explicitly says is that he ought to see to the subsistence of others at least in the cases where his own life is not imperiled. The most we can say about it is that it evinces some (I think quite natural) temptation in the direction of the Hobbesian position, while at the same time finding no basis whatever in the premises with which Locke has furnished his account.

That final point is worth stressing. Locke is often regarded as the theorist of individualism *par excellence*. As we saw, this was Fletcher's interpretation of him,¹¹³ and it resonates throughout the recent reception of his work.¹¹⁴ Even Michael Walzer, who acknowledges the "collectivist" basis of Lockean natural law, suggests that Lockean self-sacrifice "seems inconsistent with the general pattern of his political thought."¹¹⁵ Yet as one reads Locke closely, one finds *many* passages that raise this inconsistency, so many in fact that one is hard pressed to find the pattern that was supposed to give rise to his reputed individualism in the first place. His theory is materialist, and it recognizes none of the patriotic or civic motives that might lead a republican to demand sacrifice *pro patria*.¹¹⁶ But materialism is not the same as egoism.¹¹⁷ And Locke's position seems to be that the material interests of some individuals may rightfully be sacrificed in order to protect the material interests of others. Though this is morally an attractive position, and underlies the official story we tell about the limits on a defense of justification, it is worth noting, first, that Locke feels its inadequacy, and, second, that it is Hobbes, rather than Locke, who gives the more honest account of that feeling.

X

KADISH AND REGARD FOR RIGHTS

It was Sanford Kadish who started me on this investigation, and I want to conclude by locating his own discussion of the issue in relation to the theories we have been examining.

111. LOCKE, *TWO TREATISES*, *supra* note 2, at 271 (emphasis added and omitted).

112. There is a similar implication in Locke's claim that the needy man is entitled only to the *surplusage* of the rich man's goods. *See id.* at 170.

113. *See supra* notes 79-82 and accompanying text.

114. For example, in its influence on writers like Nozick.

115. WALZER, *supra* note 110, at 89.

116. Walzer notes this at *id.*

117. *See* JEREMY WALDRON, *NONSENSE UPON STILTS: BENTHAM, BURKE AND MARX ON THE RIGHTS OF MAN* 190-209 (1988), for a discussion of this.

In *Respect for Life and Regard for Rights in the Criminal Law*,¹¹⁸ Professor Kadish expresses some doubts about the agent-neutral theory of justification as it is normally presented. He usually does so in vaguely utilitarian terms: The defendant's action is justified when it serves the social good. But Professor Kadish notices that a utilitarian theory may have great difficulty with the proposition that an innocent victim is justified in killing *all* his assailants, however many there are, and that the only way out of this—that killing in self-defense is justified on deterrent grounds—proves both too much (that we may kill in retaliation, not only in self-defense) and too little (that we may not kill in self-defense when we are already certain that the killer will be convicted and punished appropriately afterwards).¹¹⁹

Kadish is convinced that the difficulty with utilitarian theory in this area is just a special case of its general difficulty: It takes insufficient account of individualized fairness and rights. The individual right to self-defense, he suggests, has an importance that is impossible to capture if we insist on merging or aggregating it in a general social calculus. In common with many critics of consequentialism, Kadish prefers to approach the matter “through the identification of moral rights, which require recognition no matter what policy is indicated by a calculus of utilities.”¹²⁰ The particular right he has in mind is “the right of every person to the law’s protection against the deadly threats of others.”¹²¹ That right, he says, has clear implications for self-defense: “For whatever uncertainty there may be about how much protection must be afforded under this right, it must at least, if it is to have any content, include maintenance of a legal liberty to resist deadly threats by all necessary means, including killing the aggressor.”¹²² It applies irrespective of the number of assailants, and irrespective, too, of their guilt or innocence, “because the justification of the victim’s defensive action does not arise from the wrongdoing of the threatener but from the right of the victim to preserve his life against a threat to it.”¹²³

How does this theory of Kadish’s stack up against the accounts we have been considering? A first point to note is that although Kadish’s account has what would normally be regarded as a Lockean flavor, it is in fact at variance with Locke’s “preservation of mankind” approach. As we

118. Kadish, *supra* note 5, at 109.

119. See Kadish, *supra* note 5, at 117.

120. *Id.* at 117-18.

121. *Id.* at 118.

122. *Id.* at 118-19. His argument for this implication is contractarian: The state is established among other things to protect us from deadly force; but there are some cases where nothing but our own self-protection will do; since it would be irrational to assume that by consenting to the social contract we had less rather than more protection than before, we must be permitted to defend ourselves in this exigency. See *id.* at 119; see also *infra* text accompanying note 131.

123. *Id.* at 119.

saw, that approach is consequentialist and seems to require each person to balance his own life neutrally against all the other lives that are stake in the situation—all being equally valuable in the eyes of God—and to act in the event for the preservation of as many as may be. True, Locke did qualify this with a rider that the agent is to preserve mankind in general “when his own Preservation comes not in competition.”¹²⁴ But as we saw at the end of Part IX of this Essay, that rider, though welcome and natural, is in fact very difficult to square with the premises of Locke’s approach. Kadish’s account squares with the rider, but for that very reason, Kadish’s account is inconsistent with Locke’s general preservationist premise.

The difference seems to be that the Kadish right is understood in something closely akin to what we have been regarding as an agent-relative justification. (And it deals with our terrorist hypothetical in almost exactly the way that Hobbes’s agent-relative account does.) It is a justification from the point of view of the agent, not from a neutral or social point of view. Kadish writes, “Surely it would be thought unfair to deny the threatened person the use of justified deadly force against his assailant, no matter what was indicated by any longrun, life preserving calculus, because it is *his* life that is at stake.”¹²⁵

The fact that this is not an agent-neutral theory is obscured somewhat by the fact that Kadish presents it as a matter of *fairness*. He also says that “since under this explanation the victim has a right to kill, *justice* requires that his action be legally justified.”¹²⁶ That may make it seem as though there is something neutral in the picture—fairness or justice—which generates the right in question. But that is an illusion. The rights themselves cannot be understood except as giving rise to agent-relative liberties. The fact that they are comprised in justice or fairness is simply a matter of the latter terms being used as general labels for whatever rights we have, whether their vindication is an agent-neutral or an agent-relative matter. Certainly Kadish wants to emphasize that these are moral rights: He is not skeptical about their objective basis. But as we saw in Part IV of this Essay,¹²⁷ there is nothing necessarily skeptical or subjectivist or amoral or nihilistic about an agent-relative theory. It is simply a distinctive type of moral theory, and it covers several highly respectable views (that of Kant, for example)¹²⁸ as well as some more dubious ones (that of Hobbes).

124. LOCKE, TWO TREATISES, *supra* note 2, at 271.

125. Kadish, *supra* note 5, at 117.

126. *Id.* at 119 (emphasis added).

127. See *supra* text accompanying notes 53-54.

128. Remember, though, that Kantian ethics is an agent-relative deontology, not an agent-relative rights-theory. It implies that I should have a special concern with the fulfillment of *my* duty, whereas Kadish’s view is that my special concerns are the subject matter of my rights. (Sometimes the former view is associated also with rights, as in Nozick’s claim that rights are correlative to agent-relative side-constraints. But that’s a separate point.) In terms of the discussion in SCHEFFLER, THE REJECTION

Something similar can be said about Kadish's insistence that the right in question is a right against the state. That does not alter its agent-relative character. Kadish believes that the authority of the state is founded in part in its offering protection to individuals against aggression from others. That idea is linked to self-defense as follows. The state is established to protect us from deadly force; this is the basis of its contact with each of us. However, there are some cases where nothing but my own self-protection will do. Since it would be irrational to assume that by consenting to the social contract I had less rather than more protection than before, I must be permitted to defend myself in this exigency.¹²⁹ Unlike Hobbes,¹³⁰ Kadish does not (I think) believe that the state has an independently justified authority to use force. Its justification, on Kadish's account, is nothing but an implication of individuals' rights to protection and defense. So there is no gap between the citizen's being justified in defending himself and its being illegitimate for the state to punish him (as there might be on Hobbes's account).

There is one other significant difference between the account Professor Kadish pursues and the account of Thomas Hobbes. Readers will recall that Hobbes adduces self-preservation as the epitome of reasonable action, and infers from this that it is unthinkable that the defense of one's life would not be justified. Kadish, by contrast, does not start from a natural right to self-preservation, but from a natural right to protection *from attack*. The difference is unexplained,¹³¹ but it emerges in Kadish's argument that a distinction must be drawn between the killing of an innocent assailant (who is attacking me) and the killing of an innocent bystander (who is not). Suppose, in our story of the terrorist from Part I of this Essay, that the hostage was given not a gun but a grenade, and he knew that if he used it for self-protection, civilians who in themselves posed no threat to him would be killed.¹³² In this sort of case, Kadish suggests, "a balance of utilities becomes determinative"¹³³ because the hostage has no rights at stake in the situation so far as the civilians are concerned. They are not attacking him, nor are they committed by social contract to laying down their lives so that he will not be attacked. If nevertheless he is justified in using the grenade, it can only be on the basis of a neutral balance of lives saved and lost.

OF CONSEQUENTIALISM, *supra* note 1, Kadish's view involves an agent-relative prerogative rather than an agent-relative requirement. Scheffler's discussion implies that it is none the worse for that!

129. See Kadish, *supra* note 5, at 119.

130. See *supra* Part VI.

131. For possible explanations, see *supra* text accompanying notes 38-39.

132. See Kadish, *supra* note 5, at 122-23. This is Professor Kadish's version of the story which I adapted from his essay, *supra* Part II.

133. *Id.* at 122.

Though these differences are significant, they are less significant I think than the fact that the explanation Kadish gives has the same general *shape* as Hobbes's explanation of the justifiability of self-defense. In each case, the simple consequentialist structure of the "lesser evil" has been subject to analysis and called into question as an account of the reasonableness of self-defense. In each case, an alternative explanation is offered and preferred, one based on individual rights, conceived roughly as Hohfeldian liberties of self-protection. The content of each right is defined in terms of the life of the person who has it, and in each case, the underlying rationale is not that these rights serve some neutral good, but that some such agent-relative liberty is to be accounted among the starting points of moral analysis. Now I can imagine that it may not be altogether pleasing to Professor Kadish to have his theory of self-defense associated with that of Thomas Hobbes in this way. Hobbes has the reputation of a philosophical *enfant terrible*, and the label of "Hobbesism" was generally regarded as an insult in the century or so following the publication of *Leviathan*.¹³⁴ It might be more comforting to be associated with the human inconsistency of John Locke's theory than with the fanatical rigor of Hobbes's.

In fact Kadish's account is more moderate than Hobbes's, inasmuch as Kadish has not committed himself, as Hobbes did, to the exclusion of everything but self-interested agent-relative rights as a basis for legal and political justification. Kadish did not deny the possibility of consequentialist, social good, or even "lesser evil" justifications; and that possibility was therefore available to him, to deal with the innocent bystander cases.¹³⁵ All Kadish denied—and here the association with Hobbes is, I think, entirely respectable—was that social utility accounts could furnish a sufficiently focused or compelling account of self-defense on its central ground. Moreover, this moderation was not purchased at the expense of consistency, as Locke's moderation appears to have been. There is nothing inconsistent in the view, implicit in Kadish's article, that morality comprises both agent-neutral and agent-relative justifications; inconsistency arises only when, as in Locke's account, one tries to pretend that the two of them can be inferred from a single source.

134. See SAMUEL I. MINTZ, *THE HUNTING OF LEVIATHAN* 21 (1962) (citing the following from a 1679 broadside entitled *An Elegie upon Mr. Thomas Hobbes of Malmesbury, lately Deceased*):

Here lies Tom Hobbes, the Bug-bear of the Nation,
Whose Death hath frighted Atheism out of Fashion.

..

Old Tom, with a Recanting Verse,
Must his odde Notions dolefully rehearse,
To new Disciples in the Devil's-Ar**.

135. However, for doubts about the general applicability of a 'lesser evil' defense, see Sanford H. Kadish, *Fifty Years of Criminal Law: An Opinionated Review*, 87 CALIF. L. REV 945, 967-68 (1999).

