Torture and the "Distributive Justice" Theory of Self-Defense: An Assessment

Whitley Kaufman

On August 1, 2002, Assistant Attorney General Jay Bybee presented to White House Counsel Alberto Gonzalez a memorandum addressing the question of what interrogation methods were prohibited by the Convention Against Torture and Other Cruel, Inhuman or Degrading Treatment or Punishment. The memorandum, widely seen as a rubber stamp for the Bush administration's use of torture, adopted an extremely narrow definition of torture, holding that "certain acts may be cruel, inhuman or degrading, but still not produce pain or suffering of the requisite intensity to fall within Section 2340A's proscription against torture." But while holding that the legal ban on torture "is limited to only the most extreme forms of physical and mental harm," the memorandum also addressed the question of when an interrogation method "might arguably cross the line" and constitute legally prohibited torture. The memorandum argued that criminal liability for such conduct might be avoided by pleading the "standard criminal law defenses of justification and self-defense," where the use of torture was necessary to "elicit information to prevent a direct and imminent threat to the United States and its citizens."

The memorandum provoked widespread outrage both domestically and internationally when it was leaked, as it was seen as permitting the president to authorize the use of torture while denying he was doing so. Especially controversial was the limiting of the definition of "torture" to cases where the pain was "equivalent in intensity to the pain accompanying serious physical injury, such as organ failure, impairment of bodily function, or even death." What has received far less attention are the defenses to a violation of the torture statute, especially the memorandum's claim that torturing a terrorist could be justified on the grounds of self-defense:

Under the current circumstances, we believe that a defendant accused of violating Section 2340A could have, in certain circumstances, grounds to properly claim the
defense of another. The threat of an impending terrorist attack threatens the lives of hundreds if not thousands of American citizens. Whether such a defense will be upheld depends on the specific context within which the interrogation decision is made. If an attack appears increasingly likely, but our intelligence services and armed forces cannot prevent it without the information from the interrogation of a specific individual, then the more likely it will appear that the conduct in question will be seen as necessary. If intelligence and other information support the conclusion that an attack is increasingly certain, then the necessity for the interrogation will be reasonable. The increasing certainty of an attack will also satisfy the imminence requirement. Finally, the fact that previous al Qaeda attacks have had as their aim the deaths of American citizens, and that evidence of other plots have had a similar goal in mind, would justify proportionality of interrogation methods designed to elicit information to prevent such deaths.⁶

**TORTURE AS SELF-DEFENSE?**

Why not simply justify the torture of terrorists on the grounds that the benefits (the savings of lives) would outweigh the costs (the temporary suffering of the terrorist)? The answer is that neither morality nor law allows the use of such wholesale consequentialist reasoning; indeed, if it did, the Bybee memorandum could have relied on the straightforward argument for torturing terrorists that the end justifies the means.⁷ Instead, the memorandum asserts a far more problematic self-defense justification for torture. It argues that the “doctrine of self-defense permits the use of force to prevent harm to another person,” and hence would apply to the case of the use of torture against a terrorist to prevent harm to civilians, where the “threat of an impending terrorist attack threatens the lives of hundreds if not thousands of American citizens,” assuming that the necessity, proportionality, and imminence requirements are met.⁸

However, the memorandum acknowledges that the use of self-defense principles to justify torturing a terrorist in order to obtain information that might save lives is far from a traditional use of the doctrine:⁹

To be sure, this situation is different from the usual self-defense justification, and, indeed, it overlaps with elements of the necessity defense. Self-defense as usually discussed involves using force against an individual who is about to conduct the attack. In the current circumstances, however, an enemy combatant in detention does not himself present a threat of harm. He is not actually carrying out the attack; rather, he has participated in the planning and preparation for the attack, or merely has knowledge of the attack through his membership in the terrorist organization.¹⁰
The memorandum refers to “leading scholarly commentators” in support of this argument, though it names only one, the legal scholar Michael Moore.\textsuperscript{11} Although no one doubts the scholarly credentials of Moore, his position on this issue is far outside the mainstream, and indeed is based on a novel theory of the foundation of self-defense, a theory that is so new and little known that it lacks even a standard name. For the purposes of this essay, I will call it the “distributive justice” (DJ) theory of self-defense. Though there are various versions of this theory, they share the same basic tenet that the use of defensive force can be justified on principles of the fair distribution of harm in situations where harm must inevitably fall on someone. The application of this theory can be illustrated by the terrorist case: Where a terrorist is responsible for an imminent danger to the public (for example, the planting of a bomb), and it is possible to avoid this danger by causing harm to the terrorist (for example, by torturing him into revealing the location of the bomb), principles of distributive justice permit inflicting harm on the guilty person to protect the innocent (call this the ticking bomb case).\textsuperscript{12} Thus, in this view, torturing the terrorist is a permissible form of self-defense.\textsuperscript{13}

The DJ theory is largely the creation of the philosopher Phillip Montague, who first articulated it in his 1981 article “Self-Defense and Choosing Among Lives.”\textsuperscript{14} Montague argues that self-defense should be understood in terms of the “distribution” of harm between persons based on the culpability of the parties in question. When a “forced choice” exists between the lives of two people, he argues, “justice permits” the nonculpable party to choose his life over that of the culpable party.\textsuperscript{15} Roughly speaking, a “forced choice” exists when a person has brought about a situation in which there is no possibility of avoiding the fact that someone must be harmed, and the only question is who should be the one to accept the harm. According to the theory, the principles of distributive justice dictate that it should be the attacker who absorbs the harm, since he is the one who culpably caused the situation in which there is a forced choice. Montague later systematically developed his theory in his book \textit{Punishment as Societal Defense};\textsuperscript{16} and his original idea has inspired many other thinkers to develop different versions of the distributive justice theory.\textsuperscript{17} As we have seen, the influence of this academic theory has now infiltrated public policy, and has been used to justify the use of torture against detainees (though of course we cannot know just how much influence the theory actually had in practice).\textsuperscript{18}
The use of DJ theory to justify torture, first proposed by Moore in 1989, has recently gained at least one influential supporter in Jeff McMahan, though it is not clear whether all advocates of the distributive justice theory would endorse its application to justify the use of torture. For the purposes of this essay, I will assume that the DJ theory would indeed justify torturing the terrorist in the ticking bomb case, in contrast to all other theories of self-defense, none of which has ever been used (to my knowledge) to justify torture. Given the lack of any legal precedent for the invocation of self-defense to justify the use of torture, or even any basis in traditional accounts of self-defense, the argument as presented in the Bybee memorandum relies entirely on the truth of the DJ theory. Hence, to assess this argument will require an inquiry into the theoretical foundations of self-defense doctrine. The approach of this paper will thus inevitably be abstract, and perhaps frustrating to those who are skeptical of the importance of abstract theory in the law. However, this very debate is arguably an answer to such skepticism, for it reflects a case where abstract theory is invoked to justify concrete public policy outcomes—that is, the use of torture. The principal goal of this paper, then, will be to demonstrate that the distributive justice account is an implausible and defective theory of self-defense, and therefore cannot be used to justify torture. Implicit in this argument is also the claim that no plausible theory of self-defense could ever justify the use of torture. If torture is ever to be morally justified, it cannot rely on a purported claim of self-defense.

**Distinctions and Definitions**

It is a notorious failing of legal and moral philosophy that we do not have an account of why self-defense is justified; that is, despite the near-universal belief that the use of force (even deadly force) in self-defense is morally and legally permissible, we lack an explanation of why this should be so. Ordinarily, both law and morality strictly prohibit the intentional killing of another human being, even if one can achieve good by doing so. To take a gruesome but famous example, in the case *R v. Dudley and Stevens*, a group of shipwreck survivors in a lifeboat had run out of food and water and chose to kill the cabin boy and eat him in order to stay alive. After the remaining survivors were rescued, they were put on trial for homicide and found guilty, despite the desperate circumstances, on the principle that it is wrong to deliberately take the life of another even to save oneself. Based on this principle, it would seem that self-defense is equally impermissible,
in that it involves deliberately taking the life of another in order to save one’s own life. Nor can one justify self-defense in that it uses force only against guilty parties—for example, by asserting that what is wrong in the Dudley case is that the cabin boy was innocent—since self-defense is permitted even against innocent aggressors (for example, a psychotic aggressor or one who mistakenly believes he is being wrongly attacked).

First, a methodological question: How do we go about assessing a theory of self-defense? I will follow here several standard practices in the literature. First, it is widely accepted that the law of homicide, and self-defense doctrine in particular, is informed and influenced by principles of morality, which help determine the legitimacy of the scope and limits of self-defense. Second, the relevance of morality to self-defense doctrine permits reliance on both moral theory (thus, for example, we have already rejected the moral theory of consequentialism) and also moral intuitions about particular cases. Third, I follow the widespread practice of considering both personal self-defense and national self-defense: that is, the principles of self-defense as found in domestic criminal law, and principles of legitimate self-defense governing nations in international law. Both are founded on the same general principles and can help inform each other. Finally, I follow the standard method of assessing a moral or legal theory of “reflective equilibrium,” which tests the adequacy of a theory by measuring it against two distinct requirements. First, the theory must successfully account for the current principles and practices of self-defense, including moral intuitions. Note that a good theory need not justify all of the present provisions of self-defense law; indeed, if the theory simply approved current law wholesale, we might be suspicious that it was a mere apology for the status quo. Still, we will insist that a plausible theory must make sense of most of current self-defense law, else it will not count as a theory of self-defense but of something different. Second, the theory must be internally coherent and consistent. It will show us how plausible underlying principles of morality and law make sense of why we permit self-defense, consistent with the ordinary presumption that intentional killing is wrongful. It must, in other words, explain to us why self-defense is justified. If the theory appears too ad hoc and arbitrary, or if the principles it appeals to are unconvincing, we cannot accept it.

The basic principles of self-defense that the theory must explain are as follows: (1) the force used must be against an Unjust Aggressor; (2) the use of force must be necessary to protect oneself; (3) the amount of force used must be
proportionate to the harm being threatened; and (4) the threat must be imminent. For the purposes of this discussion, I will grant that the proposed use of torture in the ticking bomb case satisfies the last three—necessity, proportionality, and imminence—as argued in the memorandum.24 However, faced with the first condition, the distributive justice theory runs into trouble, and the argument for torturing terrorists breaks down. Self-defense doctrine distinguishes two types of aggressors—Justified and Unjust—but importantly, the Unjust includes both the Culpable (that is, malicious) and the Innocent (not culpable). On the traditional account of self-defense, the Unjust Aggressor requirement disallows the use of force against a Justified Aggressor (for example, a policeman making a lawful arrest), but crucially it allows force against an Innocent Aggressor (for example, the insane), whose attack is not justified but is excusable, and who is thus also classified as an “Unjust Aggressor” along with the Culpable Aggressor. It will help to schematize four possible classes of people against whom force may and may not be used:

- **Culpable Aggressor.** The classic case of justified self-defense, the murderous attacker.
- **Innocent Aggressor.** For example, the axe-wielding psychotic who would be acquitted by reason of insanity; though his act is excusable, the use of self-defense is permissible despite his innocence.
- **Culpable Bystander.** The most confusing case but also the most important. It involves a person who has committed a wrongful act, but is not now committing a wrong. Because he is no longer an attacker but now merely a “bystander,” one may not use defensive force against him. For example, suppose the cabin boy in *Dudley* had earlier tried to kill one of the sailors, but had failed and now was too weak to cause trouble any longer. The remaining sailors cannot kill him on “self-defense” grounds, however culpable he may be, because he is not now an attacker but merely a bystander, and hence not an Unjust Aggressor. To kill a bystander to save oneself is not self-defense, but self-preservation; it represents what I will call instrumental force rather than defensive force: force used as a means to further a good end (to save one’s life).
- **Innocent Bystander.** The clearest case in which self-defense is *not* permissible; indeed, it is not “defense” at all because the person is not an attacker. The cabin boy in *Dudley* is such a case. The use of “human shields” to
protect oneself is an example of using Innocent Bystanders for self-protection, and is thus prohibited under the laws of war; it is expressly forbidden under the Fourth Geneva Convention.\textsuperscript{25}

There is, however, an important exception to the Innocent Bystander rule: One may cause harm to Innocent Bystanders if it is foreseen but not intended. Though the moral status of the foreseen/intended distinction is controversial, the distinction is well established in the international law of self-defense.\textsuperscript{26} Thus, the Geneva Conventions permit harm to civilians as the collateral (that is, foreseen) effect of an otherwise permissible military bombing mission, but prohibit the intentional targeting of civilians (Innocent Bystanders).\textsuperscript{27} Indeed, this is the legal and morally intuitive basis for the crucial moral distinction between terrorism (deliberately harming bystanders) and legitimate military strikes (which foresee but do not intend harm to civilians).\textsuperscript{28} More controversial is whether this distinction applies to individual self-defense. There is no doubt that the ordinary presumption is that any use of defensive force that causes harm to Innocent Bystanders will probably be reckless and unnecessary; that is the nature of the difference between the chaos of war and the relative order of peaceful society. However, there is some reason to believe that at least in some cases the foreseen harm to bystanders might be legally and morally permissible.\textsuperscript{29} Further, moral intuitions are widely seen as supporting a distinction between the two following cases:

- \textit{Trolley case}. A trolley careening out of control is heading for five people on the tracks ahead; the only way to save them is to steer the trolley onto another track. There is one person on this other track who will be killed as a result. Intuitively, it is permissible to switch the trolley to protect the five people, even though it means killing the one innocent bystander.

- \textit{Transplant case}. A doctor has five patients who will die very soon for lack of various organs for which donors are unavailable. A healthy person in his office happens to be an appropriate donor. However, the doctor may not claim other-defense in killing this one person to save his five patients. Even though this is similar to the trolley situation in that it involves the death of one to save five, the key difference is that in the transplant case the harm to the one is intended; in the trolley case it is merely foreseen. The transplant case is impermissible because it involves instrumental rather than defensive force—that is, force used as a means to save lives rather than to fend off an attack.
These, then, are the principles and intuitions regarding self-defense against which a theory of self-defense must be tested. As we will see, the distributive justice theory of self-defense fails to account for the above-described principles of self-defense and also lacks sufficient internal coherence, so that it must be rejected. As such, it cannot be used to justify the use of torture.

Putting Theory to the Test

It is relatively easy to demonstrate that the DJ theory is unable to account for the fundamental principles governing self-defense. Here we will focus on its most obvious area of failing, and the one most relevant to the present topic of torturing terrorists: its inability to make sense of the Unjust Aggressor requirement. Even if DJ theory can justify self-defense against the Culpable Aggressor, it fails to give an explanation of any of the other cases.

Innocent Aggressor
Perhaps the most obvious example of where distributive justice fails to account for the basic principles of self-defense is the Innocent Aggressor case. It is a basic principle of justice that one may not deliberately harm or punish the innocent. Yet as we have seen, the permission to harm even the innocent in self-defense is deeply embedded in our intuitions, practices, and traditions. Classic examples of Innocent Aggressors include the psychotic aggressor, the juvenile aggressor (a young child bearing a weapon), and the mistaken aggressor, who reasonably but mistakenly believes he is defending himself against an attack (this last case is quite common in warfare under the rubric of “friendly fire”). Under the doctrine of self-defense, one may use force against Innocent Aggressors, and one may in certain cases even cause indirect harm to Innocent Bystanders (the trolley case, strategic bombing), though one may never intentionally harm Innocent Bystanders or utilize human shields. But the DJ theory is unable to explain why it would ever be permissible to deliberately harm an innocent person; indeed, the theory would seem to dictate quite simply that self-defense is not permissible against innocent threats.

A common strategy among proponents of DJ theory to address this problem is to expand the notion of “culpability” as far as possible, so that many apparently Innocent Aggressors turn out to be guilty after all. Montague, for example, insists that the relevant notion of culpability for purposes of self-defense should
not be limited to cases of intentional harm. It is morally significant, he writes, that a life-threatening situation is created by any sort of culpable behavior, be it "intentional, reckless, or negligent." What is more, even the most minor of personal faults may be sufficient to subject oneself to being killed. Thus, for Montague even "excessive self-confidence" suffices to permit the distributive justice principle to take effect; he gives the example of the overconfident sailor who knew a storm was coming but decided to sail anyway, and did not bring an extra life vest for his passenger. In such a situation, it would be appropriate for the passenger to insist on using the one vest, leaving the sailor to die. So too is "vanity" sufficient to justify killing someone in self-defense: If someone has a "serious vision problem" but refuses to wear glasses out of vanity and then swallows a pacemaker by accident, it is justifiable to kill him to retrieve the pacemaker.

Other defenders of distributive justice hold that a person may be sufficiently "culpable" despite lacking any fault at all. Jeff McMahan poses the case of the man who loses control of his car because of brake failure, where he is not in any way responsible or at fault or negligent in any way. Yet it is permissible, McMahan insists, for an endangered pedestrian to kill him in self-defense, because the man "chose to drive knowing that there was a small risk that he would lose control of the vehicle and imperil the lives of others."

Some proponents of distributive justice come close to denying the very possibility of an Innocent Aggressor. Noam Zohar, for example, insists that there is always some level of guilt in an aggressor: "it is only the existence of some measure of guilt that labels an agent an 'aggressor.'" Thus, even the homicidal maniac, though he might be "acquitted in a criminal court," cannot be said to have "complete moral innocence." Similarly, a nine-year-old child waving a gun is a minor and hence not "capable of full criminal guilt," but surely he is a "bad child" and thus bears some level of culpability. And as long as there is some "minimal level of guilt" in the attacker, Zohar argues, it "tips the otherwise balanced scales of 'life versus life,'" justifying his being killed in self-defense. But surely the concept of an Innocent Aggressor is not an oxymoron; common sense tells us that there are such things as genuinely Innocent Aggressors: the insane, the mistaken, and the juvenile (as even Zohar recognizes in the case of the two-year-old).

In any case, it is highly questionable to suggest that conduct that would in a court of law be considered blameless would count as sufficiently culpable to justify execution for purposes of self-defense. In narrowing the relevant
definition of "innocence" so radically, the DJ theory undermines its claim to be a theory of justice. McMahan even suggests that responsibility for a threat can provide a basis for liability to defensive force "even in the absence of culpability or fault," as in the case of the driver who knows there is a small risk of causing harm to others. Moreover, even if one conceded for the sake of argument that there might be some degree of genuine culpability, however slight, in all (or most) aggressors, it would hardly follow that justice permits killing them on account of this minor fault. Mere negligence is indeed grounds for recovery of damages, but it is hardly sufficient to justify killing someone. Even the staunchest advocates of the DJ theory admit that justice requires that the harm inflicted must be proportional to one's degree of fault. Indeed, it is highly likely that in virtually all self-defense cases (especially Innocent Aggressor cases) there is likely to be some degree of fault on both sides. It would seem to be much more convincing to hold that justice requires extreme culpability before one forfeits one's right to life.

In any case, the best that can be said for the strategy of expanding the notion of culpability is that it may account for some cases of what are usually seen as Innocent Aggressors by redescribing them as Culpable Aggressors. But in the end the strategy merely avoids the question: Is there a right of self-defense against Innocent Aggressors? In his Punishment as Societal Defense, Montague does not tell us whether one may kill an Innocent Aggressor. He finally addressed this question in his 2000 article "Self-Defense and Innocence: Aggressors and Active Threats," admitting that "there would be nothing in self-defense situations involving [Innocent Aggressors] that could defeat the moral presumption against causing their deaths." This is a remarkable (if grudging) admission, but one which few defenders of the theory have been willing to concede. It would require a drastic and implausible departure from traditional self-defense doctrine, for it cannot be the case that law and morality require a person to stand by and let a psychotic axe murderer butcher his family rather than use force against such an Innocent Aggressor. As McMahan explains, the Innocent Aggressor "constitutes a case in which [DJ] fails to account for common intuitions about the justifiability of self-defense." Some defenders of distributive justice have suggested that we reject this intuition, and disallow self-defense against Innocent Aggressors. But few are willing to go this far, and distributive justice is left unable to account for this basic feature of self-defense.
Innocent Bystander

Distributive justice is equally unable to explain the Innocent Bystander case. Montague concedes that under the DJ theory, there is no morally significant difference between Innocent Aggressor and Innocent Bystander.\textsuperscript{42} It would then follow that so long as our legal system allows the right of self-defense against Innocent Aggressors (and there is little chance in the foreseeable future that this will change), then according to the DJ theory we are committed to allowing the use of human shields and other uses of bystanders to protect ourselves. Recall that there are cases where traditional self-defense doctrine \textit{does} permit harm against Innocent Bystanders: where it is a foreseen but not intended result of one’s action—such as the trolley problem, or in ducking an axe blow when one knows the person behind you will receive it, or in wartime the death of civilians as a by-product of a legitimate military attack. But the DJ account is singularly inept at explaining the Innocent Bystander rules.\textsuperscript{43} For the DJ theory, there would appear to be no basis for distinguishing strategic bombing from terror bombing (or trolley from transplant); since strategic bombing is permissible, then on this theory terror bombing ought to be so, too.\textsuperscript{44}

Or consider the innocent shield case, where the Unjust Aggressor uses an innocent person as a literal shield. It would certainly seem that justice would not permit killing the shield as a means to stopping the attacker, since the shield is wholly without fault or blame. Yet Moore thinks the DJ theory in fact permits killing the shield:

Consider the situation in which a terrorist (T) takes an innocent bystander (V) as a shield to use as T seeks to kill another (D). If D’s only way to defend himself against T is to shoot through V, it is common to suppose, morally and legally, that he may do so. The reason for this, I would think, lies in the “ownership” of the risk (of T’s violence) that V, through no fault of his own, has acquired. T’s taking V as a shield was V’s misfortune; V did nothing to deserve it but it’s his nonetheless, like disease or a natural catastrophe. D is entitled to resist the shifting of that risk from V to himself, not because V caused the risk, but because it’s his risk nonetheless.\textsuperscript{45}

Even if one could make sense of the rather murky notion of the “acquiring of ownership of risk,” still Moore’s argument begs the question; being an innocent shield is a misfortune, but so is being the target of T: Why should we privilege one of these misfortunes over the other? Moore seems to recognize the objection, conceding in a footnote that an “alternative account” might be to hold that
killing the shield is not justified but may be *excused*—that is, that it is *not* justified by distributive justice.

*Culpable Bystander*

The DJ theory also stumbles on the case of the Culpable Bystander (also known as the Past Aggressor)—a case that is crucial to the argument for torturing terrorists, since the captured terrorist would appear to be an example of the Culpable Bystander. Consider this particular form of the transplant case: The man who shot you and destroyed your liver has been apprehended by the police, and has the right blood type to donate his liver to you, saving your life but costing his. May you forcibly kill him and take his liver? This is, of course, the case of the Culpable Bystander. Self-defense doctrine and indeed commonsense morality and justice in general are clear: you may not do so. Self-defense is permissible only against *present aggression*, not past aggression. Even the culpability and causal responsibility of the person is not sufficient to permit forcibly taking his organs. But it is hard to see how a justice-based theory could explain this. It is certainly a forced choice, and justice would seem to require that harms be distributed so that the guilty party bears them (*he* should die from lack of a liver, not *me*). Yet this is very different from a self-defense case. As McMahan concedes, a past aggressor “is no part of the threat now, while it is occurring.” Still, such advocates of distributive justice as McMahan would apparently have to permit force against such a person, contrary to our intuitions, since on his theory anyone who has “culpably created the forced choice between himself and another person should “as a matter of justice . . . pay the price of his own culpable action.” For George Draper, the “difference between killing the immediate agent of a threat and killing a remote [past] agent of a threat . . . seem[s] to be too slight to carry the weight of the moral difference between justifiable killing and murder.”

Thus, a number of DJ advocates have suggested rejecting our intuitions and allowing the killing of the Culpable Bystander. Reem Segev, for example, hesitantly argues that the only difference between the Culpable Aggressor and the Culpable Bystander is one of timing, but timing as such “lacks apparent normative significance.” McMahan allows for killing a Culpable Bystander in his mine collapse case, where one miner is culpable of creating a collapse trapping himself and another miner, and there is only enough oxygen left for one of them. McMahan claims that, since the miner culpably “created the forced
choice” between lives, the “Justice-Based Account would imply that the other
miner would be justified in . . . killing him.” David Wasserman suggests that
the reason for this difference may be simply that the past aggressor may be re-
pentant and therefore it would be unjust to kill him. Of course, even if true, this
explanation cannot account for the unrepentant past aggressor. In any case, this
position blatantly violates self-defense doctrine, which by definition permits only
defensive force—that is, force that is used against a present aggressor to stop his
aggression.

In sum, we see that the DJ theory gives us a right to use force that is so rad-
ically different from the traditional account that it can arguably no longer even
be called self-defense. For traditional self-defense, the use of force is permissible
against Culpable Aggressors and Innocent Aggressors, but not against Culpable
Bystanders or Innocent Bystanders. For the distributive justice theory, self-
defense is permissible against Culpable Aggressors and Culpable Bystanders, but
not against Innocent Aggressors or Innocent Bystanders. Strikingly, the DJ ac-
count fully accords with traditional self-defense doctrine in only a single case: the
paradigm case of the Culpable Aggressor. For traditional self-defense, the oper-
ative principle is the fact of attack, not the fact of culpability. For distributive jus-
tice, in contrast, the operative principle is culpability, not the fact of attack. The
relevant sense of “innocence” for self-defense doctrine is not the moral notion
but the causal one: that one is a cause of harm (from the Latin nocere). But for
distributive justice, the relevant sense of “innocence” is moral in nature (though
even there it does not fit very well, if even driving a car counts as a form of culpa-
bility). In sum, the permission to use force as justified by distributive justice dif-
fers so radically from the doctrine of self-defense that distributive justice cannot
reasonably be called a theory of self-defense at all.

THE PROBLEM OF INTERNAL COHERENCE

We have seen that the distributive justice theory is unable to account for the tra-
ditional and intuitive principles governing the right of self-defense. But there is
an even deeper problem: it is doubtful whether distributive justice is even suffi-
ciently internally coherent to count as a theory at all. Recall that its essential
feature is the combining of two distinct ideas, the distributive justice principle
and the idea of a forced choice. Here we show that the idea of “justice” as used
by the theory is not a clear and coherent notion, that we are given no clear
definition of a forced choice situation, and, moreover, that we are never told why the justice principle should be limited only to such situations. In short, we are not given satisfactory definitions either of justice or forced choice, nor told how they can meaningfully fit together.

One of the most notable omissions in the distributive justice theory is its failure to get much beyond vague assertions of “justice” without presenting a theory of justice or saying how it explains self-defense. This omission is troubling, as it appears to assume far more agreement on what “justice” means or what it requires than is in fact the case. As political philosopher Jean Hampton reminds us, “There has never been, and there is not now, agreement on the nature of justice.” And there is even less agreement on the more specific doctrine of distributive justice. Libertarians, such as Robert Nozick, have questioned whether there even is such a thing as distributive justice, rejecting the presupposition that justice requires any sort of distribution or redistribution. But even those who do endorse the idea of distributive justice have traditionally limited it to the distribution of benefits: wealth, health care, jobs, property, and so on. To my knowledge, distributive justice is the first theory ever to seriously propose the distribution of violence: of physical harm, suffering, and death. It is hard to see how such a theory can be claimed to rest on widely accepted intuitions.

Indeed, it is not even clear that DJ theory is based on distributive justice, despite the vehement protestations of its advocates. It seems that DJ theorists insist their theory is based on distributive rather than retributive justice because the tradition has already explored and rejected a retributive justice theory of self-defense. But the very reason for rejecting a retributive justice theory is that it is impossible to explain the right of self-defense solely in terms of fault or culpability—precisely what the theory tries to do under the rubric of “distributive justice.” Both make the mistake of basing self-defense on the fault of the aggressor, and thus both mistakenly entail that self-defense is permissible against the Culpable Aggressor and Culpable Bystander but not the Innocent Aggressor and Innocent Bystander. Thus, to defend a fault-based theory as a form of “distributive justice” is simply to repackage the old discredited retributive-punishment theory under a new name.

Indeed, even if distributive justice were correct that moral fault is the basis for self-defense, there would be further problems for it. In discussing the theory so far we have simply assumed that the relevant fault is of one particular kind only: fault in causing the situation in which a forced choice is required—that is, being
at fault in causing the threat to others. Thus, McMahan explains, “one may not harm a culpable person, even in self-preservation, if his culpability is unrelated to the threat to oneself.” But why should we assume this? A theory of cosmic justice would presumably take into account all sorts of faults. Why should justice not apply to all sorts of forced choices, no matter how they originate? On this theory one should be able to take the vital organs of wicked people (say, convicted felons) in order to save the innocent (this is surely a “forced choice,” even if not forced by the wicked themselves), or to use wicked people as human shields. For that matter, drawing on Montague, why not use vain or overconfident people as human shields? Or consider the case where an especially wicked person uses self-defense against a mistaken but virtuous aggressor. On what theory of justice does the trivial fault in the virtuous person in causing the danger outweigh the extreme moral faults of the wicked person, thus allowing only the wicked person the right to use defensive force? Surely a genuine theory of justice would not arbitrarily limit itself to merely one type of fault.

Distributive justice is a theory of causal fault, that is, fault in causing a situation of potential harm, but it fails to explain why justice should be solely concerned with causal faults as opposed to other kinds of faults. If the theory is genuinely based on principles of justice, any such limitation is morally arbitrary. Surely justice (especially distributive justice) is not just about causation. Draper, for example, concedes that the theory should apply not merely to causing a harm, but also to failing to prevent one. This would seem to open up all kinds of problems: Would the rich person who failed to send food to the starving be liable to the use of force? But why stop there? What if the choice is “forced” not by people but by natural events? Interestingly, many defenders of the DJ theory have acknowledged that it is not just a theory of self-defense but of numerous other principles, including self-preservation and even punishment. If justice is the basis of the theory, there seems to be no limit to its implications for reshaping society. Indeed, why limit justice just to cases of forced choice? What about the principles of distributive justice where there is no urgent forced choice? Further, why limit ourselves to the distribution merely of harm? As Segev argues, the theory should apply to the distribution of benefits as well as burdens. And presumably it would demand distribution according to merit as well as fault. The implications of the theory, pursued to its logical culmination, are simply staggering.

But even if we were to agree on a definition of justice, say Montague’s “cosmic justice” principle that the virtuous should prosper and the wicked suffer,
arguably the most disturbing and morally objectionable aspect of distributive justice is its commitment to the particular *means* by which justice may be brought about: that is, by harming some in order to help others. It is generally taken as a core principle of justice that one may not intentionally use or harm one person even in order to help another. This is a foundational principle of deontology\(^59\) and also a central principle of commonsense morality, often expressed in the prohibition on using people as a means. But the DJ theory attempts to use self-defense doctrine to justify sacrificing some to benefit others, indeed even sacrificing the innocent. Segev’s response is to reject the Kantian principle of not harming one person to help another as merely an example of moral squeamishness.\(^60\) But this is simply to concede that the theory rejects the commonsense restriction on using people as means. If so, it would entail a radical revision of our entire ethics into a consequentialist theory, permitting, for example, harvesting the organs of criminals to save innocent lives. Indeed, it would follow that there are no constraints against harming the *innocent* either, in the name of the greater good—hence, one may harm Innocent Bystanders in self-defense, but also in general sacrifice the innocent to the cause of social justice.

Distributive justice, it would appear, is thus a disguised form of consequentialism, at least in the sense that it rejects the fundamental moral constraint against using people as a means (again, the crucial dividing line between traditional morality and consequentialism). To be sure, its advocates insist it is a highly constrained form of consequentialism, which applies only in “forced choice” situations, and only against culpable parties. But as discussed above, these restrictions are entirely ad hoc; if it is permissible to use people as a means to the greater good, why should this be limited to forced choices, and why to only the culpable (as Segev would say, would not this be mere “squeamishness”)?\(^61\) Indeed, self-defense has always been permitted, even against the innocent; so it would appear that if distributive justice is correct, one can use innocents as well. But the deeper problem is that neither law nor morality permit harming people as a means to a greater good.

This brings us to Moore’s argument for the permissibility of torturing terrorists to obtain the location of a hidden bomb.\(^62\) Moore argues that even though virtually all nonconsequentialist ethical theories would reject the use of torture, one can justify it on grounds of self-defense by using the DJ theory. The problem, as Moore recognizes, is that this is a case of a Culpable Bystander: “Terrorists who are captured do not now present a threat of using deadly force against their
captors or others." But distributive justice's refusal to distinguish between a Culpable Bystander and a Culpable Aggressor allows Moore to conclude that "such torture should be permissible, and on the same basis that self-defense is permissible." Moore is a self-described "threshold deontologist," so it would be open for him to argue that torturing the terrorist might be intrinsically wrong but permissible in rare cases where the magnitude of the harm were great enough—say, thousands or millions of lives were at stake. Indeed, this seems to be the commonsense view of torture. But Moore's decision to rely instead on self-defense principles to justify torture entails that the only constraint is ordinary cost-benefit analysis, rather than the much higher threshold of catastrophic harm.

The argument, however, is another illustration of the implausibility of the distributive justice theory of self-defense. To label torturing the terrorist as a form of "self-defense" is to twist the idea of self-defense beyond recognition. The terrorist is not now attacking anyone (he is helpless and in custody) and the torture being inflicted on him is not a form of "defense" against him. Moore allows that the "literal" law of self-defense would not justify this action, which is a way of saying that it is not self-defense at all. McMahan acknowledges, "Granted, [the terrorist's] own agency would not be the immediate cause of" the threatened harm. The Bybee memorandum adds: "To be sure, this situation is different from the usual self-defense justification..." But these concessions drastically underestimate the problem. The argument is a confusion between self-preservation and self-defense; the terrorist is a Culpable Bystander, not an attacker. Hence, harming him cannot count as self-defense.

But might it be argued that at least in some cases (the ticking bomb) the terrorist is not a mere bystander but a continuing aggressor, insofar as his act of violence has been set in motion but not yet completed, and he has the power to prevent it by revealing the location of the bomb? There are two ways of interpreting this argument. First, the claim might be that against a Culpable Aggressor, any sort of force is permissible so long as it helps prevent the harm that is being defended against. But this claim is patently false. The doctrine of self-defense permits only force of a certain kind: force that is defensive, that wards off or literally fends off the threatened harm. It has never been interpreted to permit using a person—even a Culpable Aggressor—as a means to escape harm—that is, to use instrumental force. Thus, Suzanne Uniacke rightly points out that one may not invoke self-defense to "forcibly use" an aggressor as a human shield.
against his coconspirators. Characterizing the captured terrorist as a “continuing aggressor” is already a highly questionable assumption, given that he is in custody and incapable of causing any more harm. (The fact that a person desires that the harm occur or even that he fails to prevent it is evidence of his present culpability but does not by itself constitute him as an aggressor.) But even if he were considered a “continuing aggressor,” it would not follow that any force against him is permissible. There would have to be evidence that the force being used is genuinely defensive, and not instrumental force used as a means to a further goal.

The argument, therefore, would have to show that this sort of force—the use of torture to get the terrorist to reveal information—was genuinely defensive. But this would be equally implausible. The use of torture to achieve a further goal, however laudable that goal, is a classic example of instrumental rather than defensive force. It is force used as a means to prevent future harm, just like using an innocent bystander as a human shield. Indeed, torture is often seen as the paradigm case of force that violates the ethical imperative not to use people as a means to achieve one’s goals. Consider, for instance, the moral philosopher Thomas Nagel’s classic discussion of the man whose friends are badly injured in a car wreck and who gets himself to a nearby isolated house. Once there he is tempted to twist a child’s arm to cause sufficient pain (that is, torture the child) to get the mother to give him her car keys to help his friends. Nagel concludes that this would not be permissible, for it is an example of violating the child’s right not to be used as a means, however morally significant the goal (saving innocent lives). That torture is not defensive but instrumental force can also be seen in that it is in practice irrelevant who one tortures so long as one achieves one’s goal; for example, one could try to get the terrorist to reveal the information by torturing his wife or child in front of him.

Indeed, advocates of the present argument for torture implicitly concede the point that torture would not be defensive force in the very fact that they turn to the DJ theory of self-defense to justify it. For if this force were defensive, then they would not need such a controversial and unorthodox theory; it would be open to them simply to argue that torture should be permissible on any plausible theory of self-defense. What distinguishes the DJ theory from other theories of self-defense is precisely that it does not distinguish Culpable Bystanders from Culpable Aggressors; that is, its theory of self-defense explicitly permits intentionally harming people as a means to self-protection (so long as they are
culpable). As we have seen, the Bybee memorandum asserts that this attempted justification "overlaps with elements of the necessity defense"—that is, with the quasi-consequentialist principle of harming one person to save others, an indirect admission that this situation does not satisfy traditional principles of self-defense.

Moore is surprisingly candid in admitting the permissibility of using the terrorist as a "means": it is permissible, Moore asserts, to torture the person as a "means to prevent the death or injury of others." Moore of course insists that, unlike consequentialism, distributive justice does not permit torturing the innocent to save lives: "Such an exception to the norm against torturing has of course no application to the torturing of the innocent." This assurance evades the basic objection: violating a fundamental moral norm does not become acceptable because it is only done to bad people. Nor is it even clear that such violations can be restricted to the guilty, if only because Moore gives us no rationale for why the consequentialist principle should be limited to the culpable. (What if distributive justice could be promoted by harming the innocent?) Moreover, if torture is justified as a form of self-defense, and self-defense is widely held to be permissible even against the innocent, how confident can we be in Moore’s assurance that such reasoning would never be used against the innocent? (Would being a "vain" person, or driving a car, be sufficient to render one subject to torture? Would a future U.S. administration claim that, in a time of war, all things are permitted?) Some DJ theorists have in fact explicitly wondered whether one may harm even the innocent on deterrent grounds.

We are left to wonder just how significant this culpability constraint really is even for Moore, as he inevitably takes us further down the slippery slope. Thus, for Moore, if others planted the bomb but the terrorist in custody "aided" them in some unspecified way, he may be tortured. What of someone who had "no hand" in the terrorist activity, but who possesses the information that could prevent the harm—for example, the mother of the terrorist? Moore considers and rejects the possibility that she might be considered "part of the threat" in that she fed and clothed the terrorists. Nonetheless, he finds a way to justify torturing her anyway. Her refusal to divulge the information renders her noninnocent: "Does she not, by that refusal, become a part of her son’s threat against which we are entitled to defend ourselves?" One wonders what constraints against torturing will be left, given the flexibility of the concept of fault in this theory. It must be insisted, however, that even if the use of torture were confined to the

TORTURE AND THE "DISTRIBUTIVE JUSTICE" THEORY OF SELF-DEFENSE
clearly guilty, it would still be wrong on deontological grounds; being guilty does not entail forfeiting one’s humanity or basic rights. It is thus highly questionable whether distributive justice can claim to be based in the notion of “justice” at all.

CONCLUSION

The distributive justice theory of self-defense must be rejected, for it can neither generate nor explain the basic principles of self-defense, nor provide plausible fundamental principles with which to unify the practice. As such, it cannot provide a basis for the justification of torture as self-defense. We can, I claim, draw an even more general conclusion from this discussion: Given that a basic tenet of self-defense is that it necessarily involves force used against an aggression, there is no basis for using force against someone fully in one’s control and custody, even if he is culpable, for he can at most be a Culpable Bystander. Thus, no plausible theory of self-defense can justify torturing detainees. The terrorist, however culpable he may be, is a bystander once he is in custody, and force used against him cannot be called “self-defense” without doing violence to the very term. Though we admittedly lack a theory of self-defense (a project I am currently at work on), when we have one it will not permit the infliction of torture as a means to the greater good. Such an action can only be based on the consequentialist theory, not on self-defense. We have already noted that the other major defense against liability for the use of torture proposed by the memorandum, the necessity defense, is also inapt, as it has never been interpreted to permit the infliction of serious harm or death on a person. The Bybee memorandum, therefore, fails to provide a satisfactory legal or moral argument to justify the use of torture, either based on self-defense or necessity.

NOTES

1 United States Code, Title 18, secs. 2340–2340A; Convention Against Torture and Other Cruel, Inhuman or Degrading Treatment or Punishment, Part I, Art. 1.
3 Ibid., p. 39.
4 Ibid.
5 Ibid., p. 1.
6 Ibid., pp. 43–44.
7 The memorandum, in fact, also makes a quasi-consequentialist argument for torture by reference to the legal doctrine of “necessity” or “choice of evils.” This doctrine allows a form of consequentialist reasoning in certain limited cases, permitting taking action otherwise unlawful in order to avoid a greater evil. However, this would appear to be a case of the exception proving the rule (against consequentialism), for the necessity doctrine is highly controversial and rarely applied. Even more important for our purposes, it has never been used (at least in the Anglo-American legal tradition) to justify killing or even harming people: one may pitch property off a ship to prevent it from sinking, but
not human beings (despite the attempt by the Model Penal Code to eliminate this restriction); that is to say, it rejects the consequentialist principle that one may use people as means to the greater good. For this reason, the necessity doctrine cannot be invoked to permit the use of torture, leaving the self-defense argument the essential matter of debate.


The memorandum’s observation that “nothing in the text, structure or history of Section 2340A precludes its application to a charge of torture” (p. 42) can only be taken as disingenuous, since the idea that self-defense might be used as a defense to torture was not addressed because it has never been taken seriously before.


The argument would also presumably apply to cases where the terrorist is even reasonably suspected of having responsibility for planting a bomb, where the torture might be calculated to find out if this is indeed the case.

It is standard in the literature to use the term “self-defense” as a generic term to cover cases of defense of self, defense of others, and even defense of property.


Ibid., p. 218.


To be sure, the use of torture would likely have gone ahead even without the support of such legal theorizing. Still, it is hard to say what might have happened had Justice Department lawyers concluded that there was no legal basis for permitting torture.

See, e.g., McMahan, “Torture, Morality, and Law” n. 4.


Moral intuitions must of course be used with caution, lest they become simply an excuse for the author asserting his own view. However, there is an increasingly large body of evidence of moral intuitions as actually measured experimentally, including, for example, on the trolley problem. Also, of course, intuitions can be inferred from actual practices and traditions—for instance, the fact that Anglo-American law of self-defense permits self-defense against Innocent Aggressors. There is, in fact, substantial evidence of widespread agreement on these intuitions, even cross-cultural agreement. See Marc Hauser, Moral Minds (New York: Harper Perennial, 2006), pp. 121-31.

See, e.g., Michael Walzer, Just and Unjust Wars, 3rd ed. (New York: Basic Books, 2000), p. 58; and Rodin, War and Self-Defense, p. 107. Indeed, the Bybee memorandum uses arguments from domestic self-defense to justify what is presumably national defense—i.e., the use of torture by the government for state purposes.
Bybee, “Memorandum,” pp. 42–43. Actually, the memorandum’s analysis of the imminence rule is questionable (p. 43), but I will ignore this complication for the purposes of this essay. For a detailed discussion of the imminence requirement, see my “Self-Defense, Immunity, and the Battered Woman,” *New Criminal Law Review* 10, no. 3 (Summer 2007). It is also noteworthy that several supporters of the distributive justice theory question the proportionality rule as well; see, e.g., Draper, “Fairness and Self-Defense,” p. 79; and McMahan, “Self-Defense and the Problem of the Innocent Attacker,” p. 259, n. 11.


The distinction between intended and foreseen harm is also essential to the definition of torture in the international Convention Against Torture, which defines torture as severe pain or suffering that is “intentionally inflicted,” as distinct from pain or suffering arising from “incidental to lawful sanctions.” Convention Against Torture and Other Cruel, Inhuman or Degrading Treatment or Punishment, art. 1.

Ibid.


See, e.g., *Smith v. State*, 419 S.E.2d 74 (Ct. App. Ga. 1992), holding that the use of legitimate self-defense that resulted in harm to a bystander did not give rise to legal liability for assault on the bystander, given the absence of any intent to harm the bystander.


Ibid., p. 37.

Ibid., p. 48.


Ibid., p. 610.

Ibid.

Ibid., p. 611.


See, e.g., Montague’s uncertainty regarding the trolley case, *Punishment as Societal Defense*, p. 46.

Farrell considers it a “very interesting question” whether one is “justified in imposing harms on admittedly innocent individuals in an effort to deter other individuals from imposing harm” (“The Justification of Deterrent Violence,” p. 316, n. 9), under the deterrence principle, which he thinks is “exactly the same principle” that justifies self-defense (p. 302). Presumably, it is also an “interesting” question for him whether one may harm an innocent bystander as a means to defend oneself.


See, e.g., Unickie, *Permissible Killing*, pp. 188–89.


Draper, “Fairness and Self-Defense,” p. 82.


Moore, Placing Blame, p. 714.

Ibid.

Ibid., p. 715.

Ibid.


Bybee, “Memorandum,” p. 44.

That defensive force must be distinguished from instrumental force (force used as a means) is a position I defend in my “Self-Defense and the Doctrine of Double Effect” (manuscript in progress). But clearly self-defense must presuppose some such distinction, for not all sorts of force are permitted even against a present Culpable Aggressor (e.g., using him as a shield or throwing him on a grenade).


Thomas Nagel, The View From Nowhere (New York: Oxford University Press, 1989), pp. 176–80. Note that it is the nature of the harm, not the fact of the child’s innocence. Nor is it relevant that the child is tortured to influence the mother; it would be just as wrong to torture the mother to get her to release the keys.

Note also that the analogy with the case of killing the kidnapper who has indicated he will kill you in the near future is not relevant, for the kidnapper who has a present intention to cause harm (even if the harm is relatively distant) is a present aggressor (indeed, he is an aggressor simply by virtue of his holding you hostage). I discuss this case in my “Self-Defense, Imminence, and the Battered Woman.”


Moore, Placing Blame, p. 715.

Ibid.

See Farrell, “The Justification of Deterrent Violence,” p. 316, n. 9. Farrell also uses self-defense to develop an effectively consequentialist theory; he finds in self-defense a basic principle of “deterrence,” i.e., the causing of harm in order to prevent greater harm.

Moore, Placing Blame, p. 716.