

Is Nuclear Deterrence Paradoxical?²*

Jeff McMahan †

It is a common claim in the moral and strategic literature on nuclear weapons that deterrence involves paradoxes. In two new books—one by Gregory Kavka, the other by John Finnis, Joseph M. Boyle, Jr., and Germain Grisez—the aspects of deterrence that are often said to give rise to moral paradoxes are discussed with greater rigor, imagination, and thoroughness than in any previous work. Indeed, these are the two best books yet written on the ethics of deterrence; they deserve and will repay careful study. They also present a curious paradox: for Kavka is a liberal who defends the practice of deterrence based on the sincere threat to engage in mass slaughter if deterrence fails, while Finnis and his two coauthors¹ are political conservatives who advocate unilateral nuclear disarmament by the West. This paradoxical mixing of views is, however, considerably easier to unravel than the paradoxes whose elements the authors discuss in their books. It is these latter paradoxes that I will examine in this review.

THE CONDITIONS OF PARADOX

The following five conditions define what Kavka calls a Special Deterrent Situation (henceforth SDS).² (1) *The necessity condition*. In order to prevent some harmful and unjust offense, an agent must threaten and conditionally intend to apply some harmful sanction in the event that the offense occurs. (2) *The sufficiency condition*. Threatening and intending to apply the sanction should the offense occur has a high probability of preventing the offense. (3) *The magnitude condition*. The amounts of harm involved in the offense and the threatened sanction are both very large. (4) *The optimality condition*. A rational consequentialist calculation would favor forming the intention to apply the sanction. (5) *The immorality condition*.

* A review of John Finnis, Joseph M. Boyle, Jr., and Germain Grisez, *Nuclear Deterrence, Morality, and Realism* (Oxford: Oxford University Press, 1987), pp. xv+429, and Gregory Kavka, *Moral Paradoxes of Nuclear Deterrence* (Cambridge: Cambridge University Press, 1987), pp. xii+243.

† I have benefited in writing this paper from comments on an earlier draft by Gregory Kavka, Robert McKim, and Nancy Gallagher Wiesler.

1. For convenience I will refer to the three coauthors collectively as “Finnis,” choosing that name simply because it is listed first on the title page. I will also refer to them collectively in the third person singular.

2. The following acronyms will be used in this paper. DAP: Disaster Avoidance Principle; MD: Minimum Deterrence; NB: Nuclear Blackmail; SDS: Special Deterrent Situation; SR: Scrupulous Retaliation; SU: Soviet Union; UR: Unscrupulous Retaliation; US: United States; WIP: Wrongful Intentions Principle.

Ethics 99 (January 1989): 407–422

© 1989 by The University of Chicago. All rights reserved. 0014-1704/89/9902-0007\$01.00

The sanction which must be threatened in order to prevent the offense is such that the agent would have conclusive moral reasons not to apply it if the offense were committed.

Kavka claims that, when coupled with certain plausible deontological principles and the claim (which he calls the "normative assumption") that it may be permissible or even obligatory to do an otherwise wrongful act if the failure to do the act would have disastrous consequences (i.e., very much worse than the consequences of doing the act), these five conditions lead to paradoxes—for example, that it may be right to form a conditional intention to act in ways that would be wrong, that a rational and morally good agent could not form such an intention, and that it might therefore be morally required of the agent that he deliberately morally corrupt himself.

If the United States (US) is in an SDS vis-à-vis the Soviet Union, and if we believe both that the normative assumption is correct and that there are deontological principles which condemn deterrence as practiced by the US, then these paradoxes will be raised by the US's present deterrent policies. The relevant deontological principles are ones which link the morality of deterrence to the morality of actually using nuclear weapons in war—hence Kavka's label, "bridge principles." The bridge principle around which Kavka and Finnis build their arguments is what they call the *Wrongful Intentions Principle* (WIP): it is wrong to intend to do what it would be wrong to do. In the absence of a principle of this sort, there would be nothing paradoxical in the idea that deterrence requires that one threaten, intend, and risk doing that which it would be wrong to do. Similarly, the paradoxes would not arise without the normative assumption. For without that assumption the relevant bridge principle would stand unchallenged and there would be no discrepancy between what one may permissibly threaten, intend, or risk and what one may permissibly do.

Conditions 1 and 4, together with the normative assumption, rule out the idea that a policy of nonpossession of nuclear weapons (henceforth simply "disarmament") would be acceptable. They also rule out deterrence based on threats which one intends not to fulfill as well as deterrence based on threats which one neither intends to fulfill nor intends not to fulfill. Condition 5 rules out the possibility that one could have an effective policy of deterrence based on sincere threats to use nuclear weapons in morally permissible ways only—what Kavka calls "Scrupulous Retaliation" (SR).

Kavka and Finnis for the most part agree that the current confrontation between the superpowers is an SDS. They both believe that the Soviet Union (SU) has or may well have aggressive intentions and that, if it does, the only hope of preventing it from dominating the world is for the US to maintain a policy of nuclear deterrence based on sincere threats of nuclear retaliation in response to aggression; and both believe that such a policy is likely to be effective in inhibiting aggression. They agree, of course, that the amounts of harm involved in Soviet domination and nuclear retaliation would be very large. Kavka believes that there are good reasons for thinking that utilitarian considerations favor forming a conditional intention to apply the sanction; Finnis argues that the relevant calculations cannot be done at all, but nevertheless gives the impression—especially in his lurid account of the Soviet threat, in which he contends that the West would immediately fall under brutal Communist domination if it were to abandon deterrence—that he would support deterrence were it not for his belief that it violates an absolute moral prohibition. Finally, both believe that

threats of countervalue attacks are necessary to sustain deterrence and that the execution of these threats would be morally wrong.

Kavka's response to the belief that we face an SDS is to accept the normative assumption, reject the applicability of the WIP to deterrent intentions, and embrace the paradoxical conclusion that it is right for the US to practice deterrence based on sincere threats to use nuclear weapons in ways that would be morally wrong. He advocates a policy of Minimum Deterrence (MD) based on countervalue targeting, with an option of retargeting the weapons for counterforce missions if deterrence fails. However, MD is to be adopted as an interim policy only, and it should be pursued in conjunction with serious efforts to achieve mutual nuclear disarmament, which is the ultimate goal. Finnis, by contrast, accepts the universal applicability of the WIP and goes on to argue that, since the destruction of cities is ruled out absolutely, and since deterrence requires that one conditionally intend the nuclear destruction of cities, deterrence itself is ruled out absolutely, even if it seems to promise better consequences than disarmament. He thus rejects the normative assumption as well as the idea that deterrence leads to paradoxes. Because he thinks there is no feasible policy of deterrence that would allow us to fulfill our duty to resist Soviet aggression without transgressing a moral absolute, he advocates disarmament, but he expects that disaster would follow if demands of morality were to be obeyed.

If the US is in an SDS, we have three options. We can reject the bridge principles altogether, adopting some form of consequentialism. Or we can, with Finnis, reject the normative assumption and accept one of the bridge principles as absolute. Or we can follow Kavka by embracing the paradoxes.³ If I were confident that the US is in an SDS, I would think that we should opt for Kavka's solution. Deterrence—or at least the sort of deterrent policy that Kavka and Finnis believe to be the only viable one—is indeed objectionable on deontological grounds. While Kavka and Finnis both focus on the objection that deterrence requires an intention to do what it would be wrong to do, what is most objectionable about deterrence is, first, that it involves a serious risk that those whom we hire to manage the policy will, with our authorization, engage in terrorist violence on an unprecedented scale by fulfilling their deterrent threats and, second, that it unjustly imposes a concomitant risk of death and injury on millions of innocent people.⁴ But, while there are thus plausible objections to deterrence that derive from the wrongness of using nuclear weapons, most of us nevertheless find Kavka's normative assumption compelling. That is, we believe that no deontological prohibitions are absolute; all are, in extreme circumstances, overridable by con-

3. Kavka's proposal to restrict the scope of the WIP is intended as a solution to the first of the three paradoxes mentioned earlier. But it leaves the other two paradoxes in place, along with others—e.g., in the case of deterrence, the fact that a person would be willing to fulfill a threat to act immorally provides a moral reason to elect that person to national political office.

4. See Jeff McMahan, "Deterrence and Deontology," *Ethics* 95 (1985): 517–36; and Douglas Lackey, "Immoral Risks," *Social Philosophy and Policy* 3 (1985): 154–75. Kavka's paradoxes still arise even if the bridge principle we use to derive the wrongness of deterrence from the wrongness of use is the principle that it is wrong to risk doing what it would be wrong to do, or the principle that it is unjust to impose risks on the innocent in order to reduce the risks to oneself. Moreover, it may seem paradoxical in itself to suppose that one could have a duty to risk deliberately doing what one believes it would be wrong to do.

siderations of consequences. The question is then whether circumstances are such that the presumption against deterrence (as Kavka and Finnis believe it has to be practiced) is in fact overridden.

Kavka seems to believe that the presumption against deterrence will be overridden if the optimality condition is satisfied. This, however, is a mistake. For a deontological prohibition to be overridden, it is not sufficient simply that a consequentialist calculation should favor the prohibited act. Rather, as the normative assumption implies, it must be the case that the consequences of obeying the prohibition would be more than just marginally worse than the consequences of overriding it. How much worse they must be depends on how strong the prohibition is. If it is strong, then the consequences would have to be very much worse. Hence Kavka's optimality condition is too weak. We are in an SDS only if a *strong optimality condition*, which reflects the demand of the normative assumption, is satisfied. In other words, the presumption against deterrence will be overridden, so that we face an SDS, only if the consequences of abandoning deterrence would be worse than the consequences of continuing to practice it by a degree at least commensurate with the strength of the deontological constraint. Thus if the consequences of abandoning deterrence would be only somewhat worse, then the normative assumption would presumably not direct us to practice deterrence in current conditions but would instead allow the deontological presumption against deterrence to stand.

My aim in the remainder of this review is to question the claim that our circumstances constitute an SDS. In particular, I will challenge the idea that the immorality and strong optimality conditions are satisfied. In the main my strategy will be modest: to counter the objections that Kavka and Finnis raise to the claim that we could escape the conditions of an SDS by adopting either SR or disarmament. A complete defense of either of these options—if it can be provided at all—is beyond the scope of this review. But Kavka's favored policy of MD, with its deliberate risk of murderous countervalue retaliation, and Finnis's despairing embrace of disarmament as the only alternative to mortal sin are both sufficiently unalluring to make even a preliminary exploration of SR and the positive aspects of disarmament worthwhile.

SCRUPULOUS RETALIATION

If SR proves feasible then the immorality condition will not be satisfied. While Kavka expresses strong doubts about the feasibility of SR, Finnis rejects the possibility altogether. Let us consider Kavka's arguments first. The first of his arguments is based on two claims. One is that SR would provide less effective deterrence than a policy that threatened greater and less discriminate destruction. The other is that SR would involve a risk of immoral retaliation, either because our scruples might vanish in the event of an attack or because our uses of nuclear weapons might result in greater and less discriminating damage than we could foresee. Hence the net effect of SR might be to increase the probability that we would use nuclear weapons in wrongful ways.

It is of course true that a risk of immoral use would remain under a policy of SR. I will return to this problem later. For now it is sufficient to note that it is implausible to suppose that SR would actually increase the probability of wrongful use. Given that advance strategic planning would call for morally legitimate uses only, the probability of wrongful use in the event of deterrence failure under SR would be very considerably lower than the probability of wrongful use

in the event of deterrence failure under a policy (henceforth UR, for “Unscrupulous Retaliation”) that actually *plans* for wrongful use. Thus, for SR to have a higher overall probability of wrongful use than UR, it would have to have a very considerably higher probability of deterrence failure. But it seems that it would not, all things considered. While the sanctions threatened by SR are less dreadful than those threatened by UR, the effectiveness of deterrence is not just a function of the awfulness of the threatened sanctions. Another factor, for example, is whether the deterrent policy itself provides incentives for attack. And it seems clear that, precisely because its threatened sanctions are less severe, SR provides fewer incentives for attack and therefore has a lower risk of preemptive attack than UR. Of course, it may have a greater risk of other types of attack. For example, it clearly has a greater risk of all-out attack. But it seems that SR would not have a higher risk of most other types of attack. Consider, for example, the threat of a limited, coercive attack. It seems that the sanction that SR might threaten in response to such an attack would be nearly as severe as any that UR could threaten *with equal credibility*. UR might threaten more severe sanctions, but the more severe the sanctions, the less likely it is that they would be imposed—in part because of the expectation that retaliation for an attack will be calibrated relative to the degree of provocation. So SR may gain back through the greater credibility of its threats at least some of the deterrent value it loses through the fact that its sanctions are often less severe.

I will return to the question of the deterrent efficacy of SR when considering Finnis’s critique. First let us examine Kavka’s second objection, which is that adversaries could exploit one’s scruples by placing important military targets in urban areas, thereby undermining the efficacy of the policy by depriving one of permissible targets. Perhaps it is true that through the relocation of its military facilities the SU could undermine SR in the long run. But the expected benefits of such an effort would be so clearly outweighed by the costs and risks that the possibility can be effectively ruled out as a cause for concern. The costs of relocation would obviously be very great, while the prospects of success would be uncertain since the entire effort could be undermined by something as simple as a change of mind on our part. And, if we were to abandon our scruples, the consequences of relocation could be disastrous for the Soviets, since any attack we might conduct would, whatever our intentions might be, involve the destruction of their cities.

While it is difficult to say what would count as scrupulous retaliation, it is safe to assume that the destruction of cities cannot be justified. This is especially obvious in the case of what Finnis calls “final retaliation,” which involves the wholesale destruction of the adversary’s society in retaliation for the destruction of one’s own, since in this case the destruction would serve no recognizable purpose whatsoever. Finnis, however, has an argument to show that deterrence either must include a war-winning capability (which involves either a disarming first-strike capability or virtually perfect defenses, both of which he rightly claims are unattainable) or must rely on sincere threats to destroy cities, both for intra-war deterrence and in final retaliation. If successful, this argument would rule out the possibility of SR. The argument begins with the claim that at every stage in the escalatory process one must be able either to deter or to repel unacceptable retaliation by one’s adversary. “For if X can neither deter nor repel Y’s retaliatory response, X cannot hope to avoid losing the war, and so can have no reason to begin or carry on the nuclear exchange” (p. 148). Since being able to repel the retaliatory response at every level is tantamount to having a war-winning capability,

the question is whether a pure counterforce policy (i.e., one that eschews attacks on cities) that falls short of being war-winning could be sufficient to deter an adversary's responses. Finnis claims that it could not. "For at every level, Y's ability to exact overwhelmingly unacceptable reprisals on X's population is (as X and Y can each foresee) very likely to deter X from carrying out an attack." Since "Y will not suffer 'unacceptable losses' " from X's counterforce retaliation, but is willing to inflict such losses on X, "the prospect of Y's retaliation 'dominates' X's pure counterforce threats. There is thus no level at which X dare carry out an attack. And so Y might well be able to coerce X's surrender without suffering any losses at all" (p. 148).

There is much that is wrong with this argument. First, it assumes a chess-player's model of rational decision making, according to which each player's initial move is dictated by calculations of the outcomes of various lengthy sequences of possible moves and countermoves. But it is implausible to suppose that those in charge of the deterrent would actually think in this way in a crisis. If they would not, then Finnis's claim that calculations predicting ultimate defeat would deter them from initial retaliation is without foundation. This becomes evident when we apply Finnis's argument to present deterrent policies. As the passage quoted above indicates, Finnis assumes that for a deterrent policy to be effective it must provide a reasonable expectation of avoiding defeat. But, since it is implausible to suppose that both sides normally have this expectation under their present policies, it seems that, by Finnis's logic, at least one side must be unable to deter the other even under present policies.⁵

Finnis's argument in fact simply begs the question by describing the damage that an unscrupulous attacker would be willing to inflict as "unacceptable," while assuming that the damage a scrupulous defender would be willing to inflict in retaliation would not be unacceptable to the attacker. What is at issue in any particular case is precisely whether the risks that an unscrupulous attacker would face in attacking a scrupulous defender would outweigh the expected benefits of aggression, which in turn depends on whether what the attacker is likely to do to the defender if the latter retaliates is regarded by the defender as worse than what he would face if he were simply to submit. In short, the key factor is how much harm the defender is willing to endure in order to avoid capitulation. It is this which will primarily determine the credibility of his retaliatory threats. Suppose, to take an extreme case, that the defender would prefer total annihilation to surrender. In that case, it would always be rational for him to fulfill his retaliatory threats, since he would prefer any amount of counterretaliatory damage to capitulation. If the attacker knows this, then whether it will be rational for him to attack will depend on whether the fruits of "victory" would be sufficient to outweigh the losses he would suffer from whatever retaliatory attacks the defender would be able to make within the limits imposed by morality.

5. The counterretaliation with which the Soviet Union threatens to respond to what would, under our present policy, be *unscrupulous* retaliation would, if anything, be worse than that with which it would respond to scrupulous retaliation. Hence if we would be deterred from retaliating under SR, then we should also be deterred under our present policy (unless, perhaps, our hope under the present policy is to be able to deter counterretaliation—a rather unrealistic hope, however, if we would have been unable to deter an initial strike).

In most actual cases, the defender will have much more at stake in the outcome of the conflict than the attacker will. So, for example, the defender will normally be willing to endure more to retain his political independence than the attacker will to attain the fruits of aggression.⁶ This suggests that the retaliatory threats of a defender will normally have considerable credibility. And, while it is difficult to say what sorts of attack might be licensed under a policy of SR, it seems clear that the sanctions threatened would not be trivial. The prospect of losing military forces has historically proven a powerful deterrent to war and, while there is no evidence for the dogma that Soviet leaders value their military forces more than they value the Soviet population, it is clear that the loss of much of the SU's military power would be regarded as a catastrophe, especially since it would increase the SU's vulnerability to attack by other adversaries, such as China. The attacker would, moreover, have more to fear than the loss of military forces. He would also have to fear the civilian casualties that would be an inevitable side-effect of even the most discriminating nuclear strikes; he would have to fear the possibility that the defender might abandon his moral scruples when under attack; and he would have to fear other possible consequences of his own unscrupulous use of nuclear weapons, such as environmental effects, domestic dissent, the mobilization of other countries made hostile by his act of aggression, and so on. All things considered, the expected costs of aggression against a scrupulous retaliator could easily be prohibitive.

Finnis, however, has a further objection to SR as I have presented it. He explicitly rejects the suggestion that the deterrent value of SR could be enhanced by exploiting an adversary's fear of the civilian casualties that would be an inevitable concomitant of even the most discriminating use of nuclear weapons. His argument is based on two claims: "In any case where those who threaten . . . are not bluffing, what they intend to do is what they threaten to do, and what they threaten to do is what they desire the other to fear from the actions they are threatening to carry out" (p. 92). Applying these claims to the case of SR, if we would want potential adversaries to fear civilian casualties of our threatened counterforce strikes, then we would threaten to cause those casualties; since our threats of counterforce strikes would not be bluffs, we would conditionally intend to cause civilian casualties; therefore our policy not only would involve wrongful intentions but would also fail to qualify as a policy of SR.

This argument is mistaken. While Finnis is right that we threaten what we wish the other side to fear from our threatened action, and while he is also clearly right in his often repeated claim that what we target does not exhaustively determine

6. During the Vietnam War, the US made the same mistake that Finnis is making when it assumed that, because it could inflict virtually limitless losses on the Vietnamese while they had only a limited capacity to harm the US, they would see that defeat was ultimately inevitable and would capitulate rather than resist. American commentators on the war still express astonishment at the amount of suffering the Vietnamese were willing to endure, attributing this to their disregard for the value of human life rather than to a commitment, of the sort that we profess to admire, to resist foreign domination ("give me liberty or give me death"). Stanley Karnow, e.g., claims that our mistake was to assume, with characteristic naïveté, that the Vietnamese shared our own deep respect for the value of life—a respect which was displayed to advantage as we inflicted on the Vietnamese the losses which they so callously accepted (see his *Vietnam: A History* [Harmondsworth: Penguin, 1984], pp. 18, 396).

what we intend to destroy, he is wrong to claim that we conditionally intend to bring about whatever we sincerely threaten to bring about.⁷ Suppose I threaten to attack some military facility, knowing that one effect of such an attack would be to kill a number of civilians. Suppose further that I want the enemy to fear the civilian casualties, since that would increase deterrence, and my purpose in making the threat is to deter the enemy. Notice that my purpose in making the threat is different from my purpose in fulfilling the threat, which would be to eliminate some military threat. If my threat fails to deter, so that I then fulfill my threat, the killing of civilians will serve no purpose of mine; it will not form part of my plan, either as an end or as means to an end, but will be incidental to my military goal. If, as Finnis claims, "an action or aspect of an action is intentional if it is a part of the plan on which one freely acts" (p. 79), then I will not intend to kill civilians when I fulfill my threat. This seems obvious, but it contradicts the implication of Finnis's view that I conditionally intend to bring about whatever I sincerely threaten to bring about. Of course, if we reject Finnis's view, we will be committed to the superficially paradoxical conclusion that one can sincerely threaten to bring about certain consequences which, in fulfilling the threat, one would not *intend* to bring about. Yet, as our example shows, it is clearly possible sincerely to threaten to bring about events that would be side-effects of one's threatened action. This is unusual, but not paradoxical.

To show that Kavka's and Finnis's objections to SR fail is not to vindicate the policy. There are various other objections to and questions about SR which must also be answered. I will conclude this section by briefly noting two further doubts. The first of these concerns the fact that SR involves the two risks noted above: first, a risk of civilian casualties as a side-effect of counterforce strikes and, second, a risk that, for one reason or another, the weapons will be used in immoral ways. Moreover, if, as seems likely, we would want potential adversaries to fear these possible consequences of our policy of SR, then we would *intend* that SR should carry these risks; for, even if the risks were inevitable, we would desire them as a means of evoking the desired fears. That the creation of these risks would be intentional is presumably of moral significance since, if it matters whether the infliction of harm is intended or unintended, then it should also matter whether the imposition of risk is intended or unintended.

Worries about the risk of incidental civilian casualties may, however, be dismissed. Since any expected harm that might be caused to civilians within the constraints laid down by the policy would, *ex hypothesi*, be justifiable, the risk of causing this harm must also be justified. The risk of immoral retaliation cannot, however, be so easily dismissed. The fact that there would be this risk does establish a presumption against SR, but whether the presumption stands or is overridden will depend on how great the risk is. As we noted earlier, there is reason to believe that the risk would be very low: strategic planning would reflect the constraints of morality and, given the difficulties and dangers of attempting to make decisions under the pressure of a nuclear attack, any actual use of nuclear weapons would almost certainly be determined by prewar planning.

The second worry about SR is whether the uses of nuclear weapons that would be morally acceptable in war would be sufficiently menacing to potential adversaries that the threat of them, if credible, would be capable of deterring

7. In replying to Finnis's objection, I follow him in assuming that it matters morally whether the harming of innocent noncombatants is intended, or foreseen but unintended.

relevant forms of aggression. Some have claimed either that there are no uses of nuclear weapons that would be morally acceptable or that the legitimate uses are so few and so limited that even a credible threat of these uses would have only marginal deterrent value. To determine whether this is correct, we need to know what the moral constraints on the use of nuclear weapons are, which depends in part on what the justification for the use of violence in war is (since the justification will determine both the scope *and limits* of permissible violence). These issues are obviously too large to address here, though I have done so elsewhere.⁸

THINKING ABOUT THE UNTHINKABLE: UNILATERAL NUCLEAR DISARMAMENT

Kavka attempts to defend his claim that the optimality condition is satisfied by presenting a consequentialist argument for MD which appeals to the *Disaster Avoidance Principle* (DAP)—the principle that, when one must choose among policies which all risk disaster, but one does not have reliable quantitative estimates of the relevant utilities and probabilities, one should choose the alternative that minimizes the probability that some disaster will occur. The intuitive point of the principle is that, when every option involves a risk of disaster, one should do what offers the best chance of securing an acceptable outcome. Kavka's argument compares MD with disarmament in terms of two possible disastrous outcomes—large-scale nuclear war (henceforth simply “war”) and Soviet domination of the world (henceforth “domination”). He claims that, even though MD has a greater probability than disarmament of leading to the worse of the two disasters (war), it has a lower probability overall of leading to disaster than disarmament, which has a high probability of leading to domination. Hence MD is favored by the DAP.

Kavka cites nine conditions under which the application of this principle is most plausible, claiming that it is reasonable to suppose that all nine are satisfied in the case of the choice between MD and disarmament. If the nine conditions are in fact satisfied, then I think that Kavka is right both that the DAP will be the most appropriate consequentialist principle to use in comparatively evaluating MD and disarmament and that it will favor the former. If, however, some of the conditions are not satisfied, then it may be the case that the DAP does not favor MD or that, even if it does, it would not be rational to be guided by the DAP in evaluating the policies in consequentialist terms. In the remainder of this article, I will argue that it is doubtful whether three of his conditions are satisfied. These conditions are as follows. *Condition 4* (henceforth C4): “The disastrous outcomes are judged to be of roughly the same order of magnitude, that is the worse disaster may be many times worse than the lesser disaster, but it is not hundreds of times worse (or more)”; *Condition 8* (C8): “The probability of the greater disaster is not thought to be very large”;⁹ *Condition 9* (C9): “The probabilities of the disasters are not thought to be very close or equal” (p. 68). If any one of these conditions is not satisfied, then it is doubtful that we should be guided by the DAP. And if C9 in particular is not satisfied, then it is not even clear that the DAP favors MD.

8. The crucial issues are identified and discussed in Jeff McMahan, *The Ethics of Killing* (Oxford: Blackwell, 1989), chaps. 4 and 5.

9. All references to probabilities should, as Kavka stipulates, be understood to refer to the period of the next thirty years.

It is important to appreciate the force of the doubts I will raise. Recall that what Kavka must show is that the strong optimality condition is met—that is, that MD is superior to disarmament by a considerable margin in consequentialist terms. Yet *each* of the three challenges that I will advance threatens the *weaker* claim that MD is superior in consequentialist terms.

Kavka thinks that in present conditions it is best to adopt the policy (MD) which seems more likely to avoid disaster, even though it has a higher probability of leading to the worst disaster. But if conditions are in fact rather different from what he thinks they are, then it may be more rational in consequentialist terms to adopt the policy (disarmament) which has the lower probability of leading to the worst disaster, even if it has a higher probability of leading to some disaster. In consequentialist terms, it will be preferable to minimize the *likelihood* of disaster if conditions are as Kavka believes—namely, that war would not be very much worse than domination (C4), that the probability of war is low under MD (C8), and that the probability of domination under disarmament is considerably higher than the probability of war under MD (C9). But it may be preferable to minimize the *magnitude* of disaster if conditions are instead that war would be very much worse than domination, the probability of war is significant under MD but substantially lower under disarmament, *or* the probability of domination under disarmament is not significantly higher than the probability of war under MD. Let us examine each of Kavka's three controversial conditions in turn.

Condition 4.—Kavka concedes that war is a worse outcome than domination, but argues that it is nevertheless rational to prefer the policy with the higher risk of the worse outcome (MD) unless war is hundreds of times worse. This condition is too strong. To see this, consider the analogous claim about probabilities that a supporter of disarmament might make—namely, that even though disarmament has a higher probability of disaster than MD, it is nevertheless rational to prefer disarmament unless the probability of domination under disarmament is hundreds of times greater than the probability of war under MD. This is quite implausible. Defenders of disarmament should concede that the greater the probability of domination under disarmament is than the probability of war under MD, the more reasonable it becomes to favor the policy which reduces the likelihood of disaster rather than the policy which reduces the likely magnitude of disaster. If domination is many times more likely under disarmament than war is under MD, that may constitute a decisive objection to disarmament. Similarly, the greater the extent to which war is worse than domination, the more important it becomes to minimize the magnitude rather than the likelihood of disaster. War need not be hundreds of times worse than domination to make it rational to favor the policy with the lower risk of war. If war is many times worse, that may be sufficient, other things being equal.

Is large-scale nuclear war many times worse than domination? While this is too large a question to be adequately answered here, several points are worth noting. First, there is a problem about comparatively evaluating war and domination, which is that each of these terms covers a range of more specific outcomes which differ greatly in utility. For example, scenarios for domination range from the extremely gruesome to the relatively benign, so that the worst scenario for domination may be worse than the least bad scenario for war, while the least bad scenario for domination is better than the worst scenario for war. Kavka's solution to the problem this poses for the ranking of the two types of outcome is to compare the average utility of the various scenarios for each outcome. Even

when we calculate values in this way it seems clear that war is many times worse than domination. Assuming that the number of scenarios for each type of outcome is roughly the same in each case, or that the difference between each scenario and the next worse is incremental, then any difference in the averages will be determined by the range of values for each outcome; and the least bad scenario for war¹⁰ is clearly considerably worse than the least bad scenario for domination, while the worst scenario for war (involving the extinction of the human species) is so much worse than the worst scenario for domination that it is virtually in a different evaluative category.¹¹ But the evaluative distance between war and domination is in fact greater than that which is revealed when we compare averages. For the method of averaging obscures the fact that, whatever the probabilities of war and domination are under the different policies, the worst scenarios for war are likelier forms of war than the worst scenarios for domination are as forms of domination. When this fact is taken into account (e.g., by weighting the utility of each scenario for the likelihood that it would be the form that the relevant type of outcome would take if it were to occur), then it becomes clearer that war should be regarded as at least many times worse than domination.

A final point: there are many ways of distinguishing or classifying the possible disastrous outcomes of our defense policies. The case for disarmament would seem stronger if, for example, we compared the two types of policy with respect to a broader range of outcomes which included extinction as a separate outcome. The fact that disarmament involves a lesser risk of extinction is perhaps its most important advantage over MD in consequentialist terms.

Condition 8.—Kavka believes that it is important for his argument that the probability of war is not high under MD—otherwise “disaster avoidance is a rather forlorn hope and its importance pales in comparison with the goal of disaster minimization” (p. 68). But another reason why it is important for him to establish C8 is that, if the probability of war under MD is *very* low, then it cannot be significantly lower under disarmament. The question whether the probability of war is significantly lower under disarmament than under MD is one which Kavka addresses only by implication (i.e., his claim that MD has a lower overall probability of disaster suggests that he believes that the probability of war is not very much lower under disarmament than under MD). But it is a question that is crucial to his argument; for, if the probability of war is considerably lower under disarmament, it may seem less rational to be guided by the DAP. If disarmament offers a considerably lower risk of the worst outcome, then it may seem the superior policy even if it has a greater overall risk of disaster.¹²

Probability judgments in this area are notoriously fallible and controversial, but it seems sensible to suppose that the probability of war over the next thirty years is relatively low under MD, though not so low that it cannot be significantly

10. Recall that here and throughout I am using the word “war” as shorthand for *large-scale* nuclear war.

11. I argue for the importance of ensuring the existence of future generations in “Nuclear Deterrence and Future Generations,” in *Nuclear Weapons and the Future of Humanity*, ed. A. Cohen and S. Lee (Totowa, N.J.: Rowman & Allanheld, 1986). My claim about the evaluation of the outcomes assumes that we exclude as a scenario for domination the situation in which domination is later followed by war (and, as a scenario for war, the barely possible situation in which war is followed by domination). Ultimately, however, these two outcomes must be taken into account in any consequentialist analysis.

12. Compare McMahan, “Nuclear Deterrence and Future Generations.”

lower under disarmament. While Kavka concedes that the probability of war is lower under disarmament, I will advance several considerations which suggest that it is significantly lower. First, a word of caution. It is tempting to think that the probability of war cannot be lower under disarmament, since the whole *point* of deterrence, and especially MD as defined by Kavka, is to prevent war, so that to claim that disarmament would have a lower probability of war seems to imply that deterrence does not work, or is counterproductive. But this is a mistake. As Finnis rightly emphasizes (pp. 66–67), the prevention of war is not the primary goal of deterrence, which is instead the preservation of a country's political sovereignty and independence. Indeed, deterrence deliberately runs a higher risk of war than there otherwise need be in order to protect political independence: it is precisely the risk of war that deters threats to independence. If the avoidance of war (or attack) were the overriding aim, that could be more readily achieved by surrendering one's independence and acceding to the demands of aggressors. So, even if deterrence is highly effective, that alone does not imply that it has a lower probability of war than disarmament.

To assess the probability of war under MD, one must have a clear idea what MD would involve. While the label "minimum deterrence" conjures up an image of an arsenal consisting of perhaps one to two hundred invulnerably based, single-warhead strategic nuclear weapons, targeted primarily on cities, Kavka's conception of MD is somewhat different. He defines MD simply as a policy "which threatens retaliation against nuclear (or other) attacks on the homeland," contrasting this with the notion of extended deterrence, which involves more ambitious deterrent objectives such as the deterrence of attacks on allies. While he claims that "a relatively small invulnerable force of weapons would be adequate for minimum deterrence" as he defines it, whether this is true depends on the circumstances (p. 8). Suppose, for example, that the following conditions were to obtain: one must deter a number of nuclear-armed adversaries, and so must be able to cover a substantial number of targets; at least some of these adversaries deploy strategic defenses; the invulnerability of one's weapons could be undermined by advances in technology—for example, in antisubmarine warfare; and one believes, as Kavka does, that one's weapons should be capable of counterforce missions. In these conditions, all of which either obtain or might obtain in the near future, MD as Kavka conceives it could require a substantial number and variety of weapons which could be perceived by adversaries as posing a threat of a first strike. In short, MD might differ from the sort of policy we have at present by less than the label might suggest, and thus might not have a substantially lower probability of war either.

Even, however, if MD as Kavka defines it would conform to the standard conception, the main routes to war that exist under the present policy would remain—for example, preemptive attack motivated by fear of the other side's arsenal, false alarm or accidental or unauthorized firing, escalation from conventional combat, and so on. These routes to war would be virtually eliminated under disarmament, for obvious reasons. Threats of war motivated by defensive concerns would be almost entirely absent under disarmament, while aggressively motivated threats would result in limited attacks rather than in war. Putting aside implausible scenarios for the malicious destruction of an entire society, it takes two sides to have a war, and disarmament eliminates one side as a participant.

Even though disarmament removes the country that practices it as a participant in war, war will still be possible if there are other countries which either have or could acquire substantial nuclear arsenals. Indeed, a case can be made for the

claim that, although disarmament by the US would reduce almost to zero the probability of a war in which the US would be directly involved (i.e., it would preclude virtually all of the scenarios for war that have any likelihood under present conditions), it might increase the probability of war among other nuclear powers. This is because US disarmament would remove a major constraint on the possibility of Soviet aggression so that, especially if the SU began to behave in an aggressive manner, other countries would feel more vulnerable to Soviet attack or blackmail and thus might attempt to compensate for the absence of the US deterrent by acquiring or augmenting their own nuclear arsenals. In these conditions of fear, uncertainty, and uncontrolled nuclear proliferation, not only might the risk of conflict be increased but the overall number of nuclear weapons might be sufficiently high that virtually any conflict could involve the large-scale use of nuclear weapons.

While this problem is undoubtedly serious, it may contain its own solution. While the main motive for proliferation would be fear of Soviet aggression, the prospect of uncontrolled proliferation would itself serve as a powerful deterrent to aggression. In other words, the SU would foresee that acts of aggression on its part would cause other countries to scramble for the protection that nuclear weapons would offer. And it would realize that, while it might be able, through preemptive attacks, to prevent some of these countries from acquiring nuclear weapons, the demands on intelligence would be too great for it to be able to deal with all potential proliferators in this way. Thus any act of aggression would carry a risk of proliferation leading to an overall decrease in Soviet security. Moreover, while the SU might be deterred from aggression by the risk of proliferation, countries that were tempted to acquire nuclear weapons might be deterred from doing so by the risk of Soviet preemption. If such a country felt that the US deterrent offered it some protection from Soviet attack, then, while the removal of that protection would increase the country's *need* for its own deterrent, it would also increase the *risks* involved in acquiring one. For, while the SU could not effectively preempt each of a number of simultaneous proliferators, it could launch effective preemptive attacks against some. (Indeed, if the risk of preemption were insufficient to deter efforts to acquire nuclear weapons, proliferation and the consequent increase in the risk of war might be prevented by the SU successfully conducting one or two limited preemptive attacks.) In sum, if I am right both that the incentives for proliferation following US disarmament would be less than is commonly supposed, and that the risks of attempting to acquire nuclear weapons would be greater than they are at present, then it does not seem that US disarmament would increase the probability of war between other countries by much, if at all.

Condition 9.—Kavka argues that, while the probability of war under MD is low, the probability of domination under disarmament is quite high,¹³ so that C9 is satisfied. Although we may concede that he is right that the likelihood of war is low under MD, I will argue that Kavka overestimates the probability of domination under disarmament, which in fact is also relatively low.¹⁴ If I am

13. See his chap. 6. Finnis is more emphatic on this point than Kavka, often writing as if domination were a certain outcome of disarmament (e.g., pp. 73, 191, 241, and 333–34).

14. An alternative way of questioning whether C9 is satisfied is to ask whether the probability of domination would be significantly higher under disarmament than under MD. If it would not be, then, given the fact that war is a likelier outcome of MD than

right that the relevant probabilities are actually rather close, then, as Kavka notes, "following a strategy of disaster avoidance rather than disaster minimization [will] seem less plausible" (p. 68). It will also be less clear that MD has a lower overall probability of disaster.¹⁵

There are roughly two ways of arguing that the probability of domination would remain relatively low under disarmament. One is to adduce evidence for the view that the Soviet Union's policies have been primarily defensively motivated and would continue to be if the US were to adopt disarmament. The other is to argue that, even if the Soviet Union harbors aggressive tendencies, these would likely be kept in check by various constraints that would continue to operate even in the absence of the US deterrent. My argument will be of the latter sort. It will focus on the constraints that would inhibit the practice of nuclear blackmail (NB) against the US—that is, the use of nuclear threats (implicit or explicit) to coerce the US either to do that which it would be morally entitled not to do or not to do that which it would be entitled to do. The focus on NB is justified by the fact that whatever would prevent the practice of NB would also serve to prevent domination; for the SU could not dominate the world without dominating the US, and it could not dominate the US except through the use of nuclear threats. Moreover, most of the constraints that would operate to inhibit NB would also serve to some extent to inhibit many lesser forms of aggression.

If the SU were to make demands backed by the threat of nuclear attack, the US could either submit or resist. I will consider what some of the consequences for the SU might be if the US were to resist. Many of these consequences might also occur even if the US were to submit. If the US chose to resist, the SU would be pressured to fulfill its threat; for the political costs of backing down—the humiliation, the loss of credibility—would be enormous. But the possible costs of fulfilling the threat by attacking the US with nuclear weapons would also be enormous. First, an unprovoked nuclear attack on the US could provoke widespread domestic dissent, leading perhaps to a crisis of legitimacy for the government. Despite popular perceptions to the contrary, the Soviet government is concerned to have the support of the Soviet people and is sensitive to popular perceptions of its own legitimacy. The unprovoked slaughter of a large number of innocent people in a country that would be known to have voluntarily renounced its own nuclear arms would be so dramatically inconsistent with Soviet ideology and domestic propaganda that it could undermine the government's standing with the people. (The same effect would probably occur, though to a lesser extent, even if the US were to submit, since it still would become known that the SU had practiced nuclear blackmail.) Similar effects could be expected throughout the SU's Eastern European domains, where revulsion could easily escalate into rebellion.

These effects would be both more likely and more severe the larger the scale of the attack on the US. So if the Soviet leadership were to decide to engage in NB, they would presumably want to make any nuclear strikes in fulfillment of

domination is, we can infer that the probability of domination under disarmament would not be significantly higher than the probability of war under MD.

15. My contention also denies that the necessity condition is satisfied. While it is true that the probability of domination is lower under MD than under disarmament, it is not the case, as the necessity condition maintains, that the probability of domination would be high in the absence of deterrence.

blackmail threats as limited as possible. (Consideration of the worldwide environmental effects of nuclear explosions would reinforce this determination.) If, however, the US were to deploy even minimally effective strategic and air defenses during the final stages of the process of disarmament, this could undermine the Soviets' aim of carefully controlling the level of destruction involved in the fulfillment of blackmail threats. For the US would then be able to defend itself against small-scale attacks, forcing the SU to launch a substantial number of weapons to guarantee that even one would get through. The resulting uncertainty (would far too many get through? would any?) would both amplify the risk of political penalties and reduce the prospects of success.

Most of the standard moral and strategic objections to strategic defense do not apply when defenses are not conjoined with offensive nuclear weapons. There are, however, two objections to the deployment of defenses in conjunction with a policy of disarmament. One is that strategic defenses might have the potential for important offensive uses. The other is that they might be regarded by adversaries as a cover for the reintroduction of offensive nuclear weapons and thus might provoke a preemptive attack. Neither of these objections would, however, have much force against a defensive system that would concentrate on intercepting attacking weapons in the later phases of their flight and would be of only limited effectiveness against a large-scale attack—which is probably the only sort of system that the US would be able to deploy in the near future.

If a disarmed US were to implement a combined program of active and passive defenses, this would strengthen its will to resist attempts at NB. And the stronger and more evident its will to resist would be, the less likely it is that it would be subjected to blackmail threats, since NB would appear more tempting to the Soviets the more likely they would be to get what they wanted by means of threats alone.

A further deterrent to NB has already been mentioned—namely, the knowledge that the use of NB against the US, especially if it were to meet with resistance, would frighten other countries that would be capable of building nuclear weapons into developing or increasing their own arsenals. As I noted earlier, the SU could address this threat by attempting to destroy emerging arsenals preemptively. This course, however, could carry grave risks of its own; and it would be unreasonable to expect that it could be more than partially effective in heading off a universal stampede for nuclear weapons. Thus the practice of NB could have the effect of creating a world in which the SU would be confronted with a host of frightened and hostile adversaries newly armed with nuclear weapons.

Perhaps the most important point is that the Soviets would know that NB would almost certainly provoke an effort at nuclear rearmament by the US itself. Both Kavka (p. 123) and Finnis (pp. 73 and 76) assume that this would not be a problem for the SU since it could physically prevent the possibility of US rearmament through nuclear attack. But to prevent the possibility of rearmament it would seem that the SU would have to destroy, *inter alia*, the US's nuclear power plants, its major physics laboratories, and its major research universities. To do this would be catastrophic for the SU itself. The destruction of the US's nuclear power plants alone, especially if it were done with nuclear weapons, would produce disastrously high levels of radiation worldwide. The global consequences would be many orders of magnitude worse than those caused by the accident at Chernobyl. The likelihood of US nuclear rearmament would therefore be an inescapable consequence of NB by the SU.

One final constraint that deserves mention is that, even if the US had totally divested itself of nuclear weapons, the SU could never be entirely sure that the US had not retained a clandestine stockpile of nuclear weapons (e.g., cruise missiles, which could be delivered to within range of the SU by quickly modified 747s if necessary). Thus the SU could never entirely discount the possibility that a blackmail threat might be met with a threat of retaliation and that the threat of retaliation might not be a bluff.

If the US were to adopt disarmament, the SU would suddenly find itself in a position of almost unprecedented security. What would be rational for it to do?¹⁶ On the one hand, the SU could act in concert with the US, which would already have demonstrated its willingness to seek peace, to find a cooperative way of ensuring an end to the threat of universal annihilation. Certainly there would be little to lose in trying. On the other hand, it could engage in NB or other forms of aggression, either to make itself master of the world or perhaps to impose a permanent peace unilaterally. The latter course would immediately put its security in jeopardy again, risking the various disastrous consequences mentioned earlier—from domestic upheaval to proliferation, US rearmament, and war. Is it really plausible to suppose that leaders of even minimal rationality would throw away this new-found security in order to pursue goals they could have little hope of achieving? If not, then we may conclude that the probability of domination under disarmament would be relatively low and that it is therefore far from clear that consequentialist considerations favor MD at all, much less that they favor it sufficiently to override the deontological presumption against forms of deterrence that risk the commission of mass murder. If this is right, then deterrence is not paradoxical. It is immoral.

16. As an exercise, one might consider what the US would be likely to do if the SU were to adopt disarmament.