Legal Theory, 5 (1999), 311–338. Printed in the United States of America

THREATS AND PREEMPTIVE PRACTICES

Claire Finkelstein*
University of California, Berkeley

I. INTRODUCTION

Sometimes it is morally permissible to prevent someone from performing a wrongful act by harming him before he has a chance to act. It is permissible, for example, to kill someone you reasonably fear is about to kill you. Under other circumstances, however, it may only be permissible to *threaten* to harm someone to discourage him from performing a wrongful act. It is not always permissible straightaway to harm him. A military commander, for example, may threaten an enemy with disproportionate retaliation, even though it would not ordinarily be permissible to carry it out.

The first sort of situation involves the infliction of harm to *prevent* a wrongful act, that is, to make it impossible for that act to occur. The second involves the use of a threat of harm to *preempt* the wrongful act, that is, to dissuade, rather than to prevent, its occurrence.[1] I shall accordingly call practices organized around the first sort of act *preventive* practices, and practices organized around the second sort *preemptive.*

Preemptive practices have a curious structure. Let us take the following features as definitional of such practices. First, the infliction of harm is not what we might call "independently" justifiable on legitimate preventive grounds—that is, it is not justifiable in and of itself to prevent the occurrence of the wrongful act. Second, the wrongfulness of the act nevertheless justifies issuing a *threat* to inflict the relevant harm. Third, there are cases in which once the threat has failed to deter the wrongful conduct, the previously prohibited infliction of harm may *become* permissible. For example, it may be permissible for the military commander to follow through with

*Rockefeller Fellow, University Center for Human Values, Princeton University (1998–99); Acting Professor of Law, University of California, Berkeley.

I wish to thank Meir Dan-Cohen, Peter Detre, Michael Finkelstein, David Gauthier, Leo Katz, Mathias Risse, Neil Siegel, and the participants of the conference on preemptive action where this paper was presented for comments on earlier drafts of the article.

1. I am aware that we sometimes speak of a person as "prevented" from doing something when he is threatened with some harm should he choose to do it, and that we also speak of a person as "preempted" if he is physically impeded from doing something. It will be convenient, however, to have separate terms for each situation, and so I have chosen to use them as indicated above.

his threat to attack if his enemy remains undeterred. This third feature of preemptive practices raises an interesting problem: Might an otherwise impermissible act become permissible because it is performed pursuant to a threat it was justifiable to issue?

Examples of preemptive practices abound in ordinary morality. They arise not only in military exploits but also in personal relationships.[2] Such practices, however, are particularly important in the law, most notably in the criminal law. Consider, for example, the police practice of using deadly force against a suspect fleeing from the commission of a felony. Under the common law version of the practice, an official or citizen was permitted to use deadly force against any suspected felon who failed to obey a command to stop, provided that the suspect had been warned of the intention to use force in the event of noncompliance.[3] The version of the right endorsed by the United States Supreme Court is more limited. It is restricted to law-enforcement officers who have probable cause to believe the suspect poses a risk of harm to others.[4] But the structure of the practice is preemptive under either version. First, the actual use of deadly force against a suspect cannot be justified on independent grounds as a legitimate preventive measure, for it is excessive to kill a person presumed innocent to prevent him from fleeing. Second, it is nevertheless legitimate for an officer to *threaten* to use deadly force in order to dissuade a suspect from fleeing. Third, it seems acceptable in at least some cases for the officer to follow through on his threat to use force should the suspect disregard his warnings. It looks, then, as though this is a practice in which the use of deadly force is rendered permissible where it otherwise would not be by the prior issuance of a legitimate threat to use it.

Alternatively, consider the rule allowing the use of force to recapture stolen property or to effectuate reentry upon land. A common formulation of the rule is that a person wishing to use force for these purposes must first issue a request that the property be returned or the land vacated. In effect, he must threaten harm before he may inflict it.[5] Arguably, the use of force under these circumstances would be excessive as an independent preventive act before the property is stolen. Does the issuance of the threat justify the use of force in this case when it would not otherwise be justifiable to use it? Stretching the framework a bit, one might also understand the right of homeowners to use deadly force in defense of habitation along the same lines. A plausible explanation for the defense is that equipping homeown-

2. In what follows I focus on practices where the infliction of harm involves the use of violence against another person. But there are nonviolent preemptive practices as well. For example, it might be permissible for a parent to deprive a child of some good in furtherance of a prior threat it was legitimate to issue, when it would not be legitimate for the parent to inflict that same deprivation in the absence of the previously issued threat.

3. *See* Joshua Dressler, UNDERSTANDING CRIMINAL LAW § 21.03 [8][2][a][i] (1995).

4. Tennessee v. Garner, 471 U.S. 1, 11 (1984).

5. Model Penal Code § 3.06(3)(a).

ers with this right constitutes an *implicit* threat against potential intruders. The actual use of force, however, outstrips the preventive justification one would ordinarily offer for it. For if the intruder poses no risk of physical harm, the use of deadly force should not be legitimate by way of prevention, and if the intruder *does* pose a risk of such harm, the use of force can be justified on self-defense grounds, making a separate rule regarding defense of habitation unnecessary.[6]

There is, however, a way of accounting for these practices that does not assign this strange justificatory efficacy to the issuance of a threat. Suppose, for example, that more suspects were apprehended under a rule that allows the police to resort to the use of deadly force than under a rule that only allows them to threaten, but not to follow through on, a failed threat.[7] Where this is the case, it is tempting to explain the preemptive practice in collective terms, namely as a way of discouraging similar wrongful acts by other agents. The use of force would then be justifiable in the aggregate, even if it could not be justified against any particular wrongdoer. Under an aggregative rationale, the practices I am calling "preemptive" would be just another species of preventive practice. The difference between a preventive practice like self-defense and a practice like that involving the use of deadly force against fleeing suspects would thus lie in the fact that in the former the use of force could be justified in each particular case, whereas in the latter the justification for the use of force would make reference to the more systemic benefits the practice as a whole engenders.

But while the turn to aggregative justification is a natural one to make, I do not think we ought to accept it. For such a justification seems to presuppose the legitimacy of *using* a person to serve as an example for others, without needing to justify the treatment of that person on its own terms. It would sanction shooting a suspect, for example, merely as a way of deterring future suspects from flight, rather than because it was justifiable to do so in *his* case. In a liberal regime, concern with individual rights leads to an insistence on a justification for the infliction of violence that is *individualized,* namely that justifies the use of violence on a case-by-case basis. If we accept this limitation on the form of justification required to legitimate violent practices, our question about preemptive practices is squarely posed: Can acts of violence in a practice involving the issuance of deterrent threats meet the requirement of *individualized* justification, when

6. Sometimes the preemptive nature of the practice is made explicit, as when a homeowner is obligated to warn potential intruders of the presence of a protective device.

7. The majority in *Tennessee v. Garner* dismissed this as a ground for retaining the common law rule, saying that there was little evidence that appreciably greater numbers of felons were apprehended under it. Tennessee v. Garner, 471 U.S. at 10–11. But it seems likely that this provided at least part of the historical rationale for the rule, whether adequately empirically supported or not.

there is no *independent* justification that would vindicate such acts as valuable in their own right?

In what follows, I shall explore one way of supplying an affirmative answer to this question. The solution to the problem of deterrent threats I shall explore exploits a parallel between the problem of the justifiability of deterrent threats and that of their rationality. I shall thus hope to draw on reflections about the rationality of deterrent threats to suggest a possible account of their justifiability. As we shall see, one prominent theory of the rationality of *assurances,* that introduced by David Gauthier, suggests at least the outlines of an answer to the rationality of threats. A problem with exploiting this suggestion, however, is that there appears to be an asymmetry between assurances and threats that may make the account of the rationality of the former inapplicable to the latter. The central task of this article is to explore that asymmetry. In particular, I shall propose a way of understanding deterrent threats that makes Gauthier's account of the rationality of assurances applicable to them. The account of threats I present will not appeal to someone who rejects Gauthier's account of assurances. That is a drawback, in light of the fact that the latter is controversial. But it should prove interesting to see how at least one prominent theory of the rationality of assurances can ultimately provide a way of thinking about the justifiability of legal practices involving the use of deterrent threats. In addition, if I am correct that the rationality of assurances and the rationality of threats must be taken together, this would have important implications for Gauthier's recent thoughts about the relation between assurances and threats. My argument that threats can be assimilated to assurances on Gauthier's account should thus be of interest apart from its application to the problem of justifying preemptive practices.

II. THE JUSTIFIABILITY OF DETERRENT THREATS

As we have seen, preemptive practices raise the question of whether it is possible to justify an act otherwise unjustifiable by the fact that it constitutes the execution of a failed deterrent threat it was legitimate to issue. Many authors have rejected the possibility of an account with such a structure in the context of nuclear deterrence. Their central claim is that it is not legitimate to threaten to use nuclear weapons in retaliation for another's nuclear attack, because it could never be legitimate actually to use them.[8] This follows from a principle they call the *Wrongful Intentions Principle* (WIP), which says it is wrong to form the intention to do something it would be wrong actually to do.[9] Because a deterrent threat is a species of intention,

8. *See, e.g.,* Michael Dummett, *The Morality of Deterrence,* 22 CAN. J. PHIL. 111 (1986).

9. The principle was first explicitly formulated by Gregory Kavka. *See* Gregory Kavka, *Some Paradoxes of Deterrence,* 75 J. PHIL. 285 (1978).

adherents of WIP would argue that a threat could never justify an act otherwise unjustifiable. For it is only permissible to threaten those acts it would be permissible on independent grounds to perform anyway. An adherent of WIP would accordingly reject the legitimacy of any practice of the sort I am calling "preemptive."

In order to justify preemptive practices, then, we would have to think about justification differently. Let us begin by considering an account that rejects the requirement of independent justification, such as that offered by Larry Alexander of practices like self-defense and defense of habitation.[10] Alexander argues that two rather minimal principles are adequate to account for legitimate prevention in either context: the *wrongful act* principle, which requires that the act sought to be prevented is wrongful, and the *notice* principle, which requires that the potential wrongdoer be placed on notice of the harm that will befall him should he perform the wrongful act.[11] He begins with a broad view of threat-issuance: Any threat of force is legitimate to discourage the wrongful act of another. Alexander then attempts to move from legitimate threat-issuance to legitimate threat-execution by appealing to the possibility of automatic punishment. He argues that if it were legitimate to threaten to harm someone contemplating a wrongful act, it would be legitimate to program a machine to inflict harm under these same circumstances. He then claims that if it were legitimate to program a machine to inflict harm in the event of a failed threat, it would also be legitimate to inflict such harm "manually," as long as the agent performing the wrongful act had been notified of the conditional intention to inflict harm.

In a similar vein, Warren Quinn has offered an account of punishment that also rejects the requirement of independent justification. On Quinn's account, the right to punish follows from the right to threaten punishment in order to deter wrongful acts of aggression. The right to threaten punishment, in turn, derives from the right to self-defense. Like Alexander, Quinn accepts that the sole justification for the infliction of harm may sometimes consist in the fact that its infliction follows from a failed deterrent threat it is legitimate to issue. And like Alexander again, Quinn argues for the move from legitimate threat-issuance to legitimate threat-execution by imagining a device of automatic punishment, suggesting that if it were legitimate to issue a deterrent threat in self-defense, it would be legitimate to program a

10. *See* Larry Alexander, *The Doomsday Machine: Proportionality, Punishment and Prevention,* 63 THE MONIST 199 (1980).

11. *Id.* at 213. Curiously, these two rather weak principles turn out to be too strong for ordinary preventive practices like self-defense. There is no reason to think, for example, that one is under any obligation to place the attacker on notice before defending oneself against his imminent attack. By contrast, the account is more helpful for understanding preemptive practices like the fleeing felon rule, as the notice requirement makes more sense in this context. Notice is at least a *logical* condition for the issuance of justifiable deterrent threats, because a threat cannot be effective in the absence of clear notice of the consequences of violating the conditions of the threat.

device to make good on such threats. He then argues that a threat it would be legitimate to execute automatically could be justifiably executed "manually."[12]

Alexander and Quinn both accept that it is possible to justify an act of violence that is otherwise unjustifiable by the fact that it represents the execution of a deterrent threat it was justifiable to issue. The sort of account they propose thus has the right structure to explain preemptive practices. The disadvantage of the account, however, is that without some restrictions on the acts it is possible to justify in this way, we could bootstrap ourselves into justifying *any* act the threat of which was issued to deter wrongful conduct. Thus the account would justify not only the law-enforcement practices in which we are interested, but many extreme preemptive measures we would *not* wish to justify. As Alexander himself points out, the account would justify a town's setting up a machine gun to automatically fire on those trespassing on a courthouse lawn, as long as the use of the machine gun were adequately publicized to potential trespassers in advance.[13] Let us accordingly call an account of the above sort the Extreme Account, because of the extreme results it purports to justify.[14]

What we seem to require, then, is bootstrapping of the sort the Extreme Account envisages, but not quite as *much* bootstrapping as it might allow. In particular, there are two ways the account might be tempered. First, there might be limitations on the threats it is legitimate to *issue* to deter wrongful conduct. Issuing a threat, after all, is a form of coercion, and there are restrictions on the legitimacy of any sort of coercion, even when imposed to deter wrongful conduct. Is it justifiable to threaten a child who refuses to go to bed with a severe beating, the question of actually inflicting the beating to one side? Arguably not. It is not clear to me, however, how one would determine which threats are legitimate and which are not. I will thus leave this avenue unexplored.

Second, there might be limitations on the threats it is legitimate to *execute,* even assuming a broad account of the threats it is legitimate to *issue.* In particular, one might wish to question the move from threat-issuance to threat-execution via the use of an automatic retaliation device. If threat-

12. Warren Quinn, *The Right to Threaten and the Right to Punish, in* MORALITY AND ACTION 52 (1993).

13. Alexander, *supra* note 10, at 213.

14. Alexander explicitly endorses an unlimited account of threat-issuance where the threat is issued to deter wrongful conduct. He is consistent, however, and simply accepts the extreme results of his account. Quinn does not assume that every threat issued to deter wrongful conduct is legitimate. For he maintains that the right to threaten punishment depends on the right to self-defense, suggesting that he thinks there *are* limitations on legitimate threat-issuance. But the right of threat-issuance Quinn supposes will arguably still be too broad. For if the right to threaten harm is as broad as the right to self-defense, and the right to punish is as broad as the right to threaten, then the right to punish some instance of wrongdoing will be as broad as the right to prevent it. But rights of prevention are broader than rights of punishment. For example, it is permissible to kill someone to prevent him from raping you, but death is too severe a punishment for rape.

execution is automatic, rather than manual, the entire course of conduct—threaten harm under certain conditions and carry it out—can be decided upon in a single, up-front choice. Under these circumstances, there would be an independent justification for executing the threat after all, namely a preventive rationale that would apply to the decision to set up the machine in the first place. By assuming we can move from automatic to manual threat-execution, Alexander and Quinn have arguably assumed away the most important aspect of our problem, namely the question of whether there is a way of justifying an individual act of violence when there is no independent justification for performing it.

In seeking to develop an account of preemptive practices, then, let us assume that no automatic device such as Alexander and Quinn use to link threat-issuance and threat-execution is available. Let us instead assume that every deterrent threat must be what I shall call "deliberatively executed," that is, the execution of a threat must take place pursuant to a separate act of choice. In this way we can ensure that any threat justifiably issued is truly justifiably executed, in cases in which the threat has failed to deter.

Let us also add a second assumption, one that will probably seem more questionable. This assumption is that it is not possible to issue a bluffing threat. Granted, people resort to bluffs all the time, often successfully. But we can justify this assumption in the present context on the grounds that threats issued as part of a large-scale social practice would be transparent over time to their recipients. While an isolated bluff might have deterrent efficacy, a law-enforcement practice like the one we are considering would quickly become ineffective if the threats it employed were not sincere. There is also a conceptual argument against bluffing threats we will consider when we discuss their rationality. We need no conceptual argument here, however, given that our discussion of deterrent threats is meant to apply to practices of the aforementioned sort.

The no-bluffs assumption crucially transforms our problem. For it links the issuance of a threat and its execution in a way that requires us to justify the execution of any threat whose issuance we wish to justify. We can accordingly treat the legitimacy of executing deterrent threats as supplying the conditions under which deterrent threats may be issued (and vice versa).[15] We could have reached this same result by assuming that any threat justifiably issued must also be independently justifiable to execute, as WIP would require. Instead, we have tied the justifiability of threat-issuance to that of threat-execution without restricting legitimate threat-execution to those acts that admit of independent justification.

15. For this reason, we can leave the suggestion of some authors that the justifiability of threat-issuance and that of threat-execution admit of entirely separate analysis to one side. *See, e.g.*, David Lewis, *Devil's Bargains and the Real World, in* MACLEAN, THE SECURITY GAMBLE (1984). Although the suggestion could be correct in theory, it cannot be practically engaged for the sort of practice I have in mind.

Having suggested parameters for a theory of the moral justifiability of deterrent threats, I wish to turn to the problem of accounting for their rationality. As I have suggested, there is an answer to the question of when it is rational to issue and make good on deterrent threats that will provide, by way of analogy, an answer to the question of when such threats are justifiable. While the account of the rationality of deterrent threats will be significantly more precise and easier to apply than the account of their morality, it will be instructive to consider whether the vindication of their rationality provides at least a framework within which we can attempt to solve the moral difficulties preemptive practices raise.

III. RATIONAL ASSURANCES, RATIONAL THREATS

Let us abstract from the moral issues by considering an example of a clearly unjustifiable threat someone might issue. The question we will now consider is whether it would be *rational* to issue, and if need be to execute, such a threat. Suppose Abigail wishes to deter Burt from applying for a job Abigail herself wants to obtain. Abigail therefore threatens to tell Burt's wife that Burt is having an affair if Burt applies for the job. Let us suppose that Abigail will succeed in deterring Burt if Abigail can make her threat credible. Let us also suppose, however, that it will be costly for Abigail actually to inform Burt's wife.[16] (Abigail herself is the person with whom Burt is having the affair, and she has reasons of her own for wishing to keep it secret.) The trouble, then, is that it would not be rational for Abigail to make good on her threat if Burt ignores the threat and applies for the job anyway. And assuming Burt *knows* this, Abigail cannot make her threat credible to Burt.

The question, then, is whether Abigail can find some way of making her threat credible. One strategy would be for her to *precommit* to telling Burt's wife in the event that Burt applies for the job. For example, Abigail might pay someone to tell Burt's spouse should Burt apply for the job. Or Abigail might arrange matters such that Burt's applying for the job would somehow *trigger* the informing of Burt's spouse. But let us now carry forward the two assumptions we developed in discussing the justifiability of deterrent threats to our problem here. First, let us assume that no threat can be automatically executed. In the theory of rational choice, the point is often put by saying that no precommitment is possible. This assumption turns out to be quite crucial, for without it, executing the threat would be like swallowing a pill: A person might manage, by finding the right pill, to make herself behave

16. I shall follow the practice of some writers of defining a threat as a conditional intention that will be costly for the threatener to execute should the threat fail to deter. Gauthier writes that "a threat (if sincere) must be supposed by its issuer to commit her conditionally to a retaliatory act that would make the threatened party's life go less well than if he were not to incur it, and her own life to go less well than if she were not to execute it." *See* David Gauthier, *Assure and Threaten,* 104 ETHICS 690, 710 (1994).

in a way that would best serve her interests, but this would not be sufficient to defend the rationality of the actions performed under the pill's influence. The only action whose *rationality* could be defended would be the act of taking the pill in the first place. Demonstration of the rationality of a first act that causes a second act, however, does not demonstrate the separate rationality of the second act, even if the latter is one it is in the agent's interest to perform.

It is worth noting that if one were to insist on the independent rationality of any act performed in execution of a deterrent threat (the rational analogue of WIP), this first assumption would make *all* deterrent threats irrational to execute. Acceptance of a strict expected-utility account of rational action, for example, would eliminate the possibility of accounting for the rationality of issuing and making good on deterrent threats. For by hypothesis it will never be rational to choose to make good on a failed deterrent threat, since doing so will be all cost and no gain. If we now carry over the no-bluffs assumption from our previous discussion as well, we can see that were we to require independent vindication of the rationality of each act, we would also be unable to account for the rationality of *issuing* any deterrent threat. For if bluffing is ruled out, then any argument for the rationality of issuing a threat must also provide an argument for making good on the threat should it fail to deter. And given that it can never be rational to make good on a failed deterrent threat, assuming no precommitment, it can never be rational to issue such a threat either.

In the previous section we made use of a pragmatic argument for the no-bluffs assumption. Here I wish to suggest the outlines of a conceptual argument for it. Although there are pragmatic arguments to which we might turn in this context as well, if the account of the rationality of deterrent threats I shall offer is to have validity outside of its particular application here, it will be necessary to justify the no-bluffs assumption on nonempirical grounds. Assume that Abigail and Burt are rational, and that each knows the other is rational. Assume, that is, that there is common knowledge of rationality between them. Now there is an argument for the irrationality of issuing bluffing threats that proceeds by *reductio*. Suppose it *were* rational for Abigail to issue a bluffing threat. Then if both Abigail and Burt are equally rational, Burt would also know it was rational for Abigail to issue a bluffing threat. Moreover, because there is common knowledge of rationality, Burt knows that Abigail knows it is rational for her to issue a bluffing threat. Under these circumstances, no threat Abigail could issue would be credible, as Burt would suppose Abigail was bluffing. To make a threat effective it must be credible, and to make it credible, it must be the case that it is not rational to issue a bluffing threat, on pain of making it rational for both parties to accept the ineffectiveness of any threat. While the advocate of the rationality of bluffs has responses she might make, I shall not consider these responses here. Anyone disinclined to accept it can

always limit my argument to cases in which the no-bluffs assumption can be pragmatically defended.

Against the background of the above assumptions, Abigail appears to be in a bind. For she knows she will have occasion to reconsider the rationality of making good on her threat to inform Burt's spouse if Burt proceeds to apply for the job. And when she reconsiders, she will realize she has nothing to gain and something to lose by actually telling Burt's wife. She cannot appeal to the deterrent benefits of issuing and making good on her threat at *this* point, for her efforts to deter Burt will already have failed. Abigail might think it rational to issue the threat anyway, hoping to deter Burt but planning not to make good on the threat should it fail to deter. But Abigail knows that it is not rational for her to issue a bluffing threat, and she knows that Burt knows this. Given our two conditions, namely that Abigail must execute any threat deliberatively and that she cannot issue a bluffing threat, Abigail cannot regard the *issuance* of a deterrent threat as rational, for she knows she will not regard the threat as rational to execute.

Let us now turn to David Gauthier's argument for the rationality of issuing and making good on *assurances*. As I have suggested, that account may give us a way of thinking about the rationality of deterrent threats. In particular, let us consider the rationality of issuing an assurance that it would not be rational on independent grounds to perform in order to consider the parallel to the threats problem we have been discussing.

Consider the familiar example of the two farmers, each of whom needs the other to help with plowing, and neither of whom expects to require the services of the other in any future year. (Assume both are retiring from farming with the next harvest season.) Farmer Clarence's field is ready for plowing this week, and farmer Doreen's field will be ready next week. Is it rational for Clarence and Doreen to enter into an agreement to plow one another's fields? Doreen should surely conclude, if she is rational, that it would *not* be in her interest to enter into such an agreement. For although she cannot plow her field by herself, it is worse for her to help Clarence plow his field and end up with no help next week plowing her own than for neither farmer to help the other. Because Clarence has no reason to help Doreen once Clarence's field is plowed, Clarence will surely fail to make good on his promise to help Doreen. Although it would be in Clarence's interest to convince Doreen that his promise to reciprocate is sincere, since otherwise he will not gain the benefit of Doreen's assistance, it seems impossible for him to offer the assurance necessary to convince Doreen to perform first.

Here, however, is an argument for why it *would* be rational for Clarence to make good on a promise to assist Doreen next week. If Clarence could find someone who would be willing to force him to assist Doreen next week, say, for a fee of $10, it would be rational for Clarence to spend the $10 to secure Doreen's help now. Would it not then be rational for Clarence simply to commit mentally to help Doreen next week? For not only could he then secure

Doreen's help in plowing his field this week, he could do so without spending the $10 he would otherwise have to spend on precommitment. If precommitment is rational, then internal commitment must be all the more rational, and Clarence and Doreen should both regard the rational solution as that which involves Clarence's prior sincere commitment to assist Doreen.

But what is to stop Clarence from reconsidering his intention to help Doreen once next week rolls around? And if Clarence *knows* he will have a chance to reconsider his intention, what can he tell himself to make his commitment to helping Doreen a sincere one *now*? Because Clarence cannot make use of any actual precommitment, what he needs is an argument he can give himself for making good on his assurance when the time comes to do so. That is, he needs a *deliberative* principle for vindicating the rationality of performing an act which is suboptimal considered on independent grounds.

Gauthier argues in favor of just such a deliberative principle. For he claims that there is a form of reasoning Clarence can use to reap the benefits of precommitment without incurring its costs. The deliberative principle can be stated in the form of a recommendation: Adopt that course of action associated with the best outcome from among the various courses of action available, and then settle on particular actions by determining what executing that course of action requires. The agent who has adopted a plan must not deliberate about what actions to perform by recalculating the independent benefits of each action separately. This recommendation applies, at any rate, if it continues to be the case that the agent could expect to do better under the selected plan than under any other available plan. He should therefore act in accordance with any plan that was rationally adopted, as long as no change in his situation occurs that would make that plan suboptimal. The required deliberative principle that allows us to move from the issuance of an assurance to its performance in the absence of precommitment, then, is a principle about the optimal form of reasoning itself. It advises an agent to reason in two tiers: optimize over plans, and select actions in accordance with the plans thus adopted.

This account of the rationality of assurances seems to solve the problem of the rationality of deterrent threats. For it suggests precisely the sort of principle we need to move from the rationality of threat-issuance to the rationality of threat-execution in the absence of automatic retaliation. When Abigail attempts to deter Burt from applying for the job Abigail covets by threatening to tell Burt's wife of the affair, Abigail adopts the plan it is most rational for her to adopt. When the time comes to decide whether to make good on the threat, Abigail should not deliberate about the benefits of informing Burt's wife in isolation. Instead, Abigail should reason about that action by placing it in the context of the previously adopted plan. And since the plan calls for Abigail to make good on her threat where Abigail has failed to deter Burt by issuing it, Abigail should act as her plan requires and tell Burt's wife about the affair.

Let us skip ahead briefly to the use I wish to make of the rational choice argument and see how the analogous account of the justifiability of deterrent threats would go. It would be justifiable to issue and make good on a deterrent threat when the action in fulfillment of the threat follows from a morally justified plan taken as a whole. While it may not be justifiable to use deadly force against those merely suspected of having committed a violent felony on independent grounds, the entire course of conduct—threaten to use force if the suspect continues to flee and use force in the absence of compliance with the threat—may be justifiable. Rather than reasoning *directly* about whether to execute a particular failed deterrent threat, a moral agent should reason about the justifiability of such actions indirectly, relativizing them to plans: If the plan of threat-issuance and threat-execution produces a morally better state of affairs than would obtain without the plan, the use of preventive force would be justified if required by the plan.

In the theory of rationality, reasoning by reference to plans rather than individual acts is vindicated under what is sometimes called a *pragmatic* theory of rationality.[17] Such a theory considers how the individual reasoner can make herself as well off as possible, measured solely in terms of possible outcomes. A question for the foregoing account of the justifiability of deterrent threats concerns the analogue of the pragmatic theory of rationality on the moral side: What *moral* theory could we use to judge the justifiability of overall courses of action? Here I wish merely to note the difficulty. I shall return to it briefly at the end.

I have now sketched the argument in its entirety. We must, however, return to address a rather significant objection. I proceeded by applying an argument for the rationality of following through on assurances to the rationality of following through on deterrent threats. I then applied the argument for the rationality of deterrent threats to the moral justifiability of issuing and making good on such threats. But the first step in the argument is highly problematic, insofar as there seems to be an asymmetry between assurances and threats that may make the argument for the rationality of the former inapplicable to the latter. I shall devote most of the remaining discussion to an exploration of this supposed asymmetry. I conclude that there is not in fact the asymmetry between threats and assurances there appears to be. To the extent an asymmetry exists, it is instead between plans involving individually suboptimal actions whose payoffs are risky and those whose payoffs are sure. This asymmetry does not differentiate the rationality of plans involving assurances from those involving threats. As we shall see, it divides the terrain somewhat differently. If this is correct, the argument for the rationality of assurances we have considered will apply to threats after all.

17. *See* Edward F. McClennen, RATIONALITY AND DYNAMIC CHOICE: FOUNDATIONAL EXPLORATIONS § 5.2 (1990).

IV. AN ASYMMETRY BETWEEN THREATS AND ASSURANCES?

Let us suppose it is possible to defend the rationality of performing a suboptimal act in satisfaction of an assurance as suggested, namely by reference to the fact that the act is part of a plan it was optimal to adopt. The problem is that where threats are concerned, I may not be able to reason in this way. For once my threat has failed to deter, I will regard the plan of attempting to deter you by issuing a threat as a poor plan to have adopted. *Now* I see I would have been better off never having adopted the plan of trying to deter you in the first place, for I have gained nothing by doing so, and the plan will now be costly to execute. An apparent difference between threats and assurances, then, is that I will regret having adopted the plan if I end up having to execute my threat, whereas in the case of a plan involving an assurance I will remain pleased with the plan, even when I must incur the cost of making good on the assurance. This seems to suggest that unlike where assurances are concerned, it surely cannot be rational for me to adopt a plan involving the use of a deterrent threat. For the conditions under which I would have to execute the threat are ones under which the plan will turn out to have been disadvatageous to adopt.

In his article *Assure and Threaten,* Gauthier attempts to solve this problem by suggesting that an agent might regard herself as advantaged by a *policy* of issuing and making good on threats, even if she cannot regard herself as advantaged by any particular deterrent threat. The difference between assurances and threats, he suggests, is that the policy does the work for threats that a course of action does for assurances: Where threats are concerned, an agent must optimize over policies, and then decide on courses of action involving deterrent threats by reference to the policy. Thus, the argument for executing a failed threat is that issuing and executing deterrent threats is part of an overall policy it is optimal to adopt, given the deterrent benefits of maintaining such a policy.[18]

Gauthier now acknowledges, however, that the appeal to policy does not work. Briefly, the objection that troubles him the most is an argument to the effect that it can never be rational to *institute* a policy of issuing and making good on deterrent threats, and so it cannot be rational to stick to such a policy when the policy requires the performance of a suboptimal act.[19] For suppose an agent issues a first threat as part of an intended policy of issuing and making good on threats, and suppose *that* threat fails to yield any deterrent benefits. She can only vindicate the rationality of making good on the threat by appealing to possible *future* benefits of the policy. But what is to guarantee that such future benefits will indeed accrue? Granted, if she happens to find herself in the middle of such a policy she can

18. Gauthier, *supra* note 16.

19. This objection is set out by Joe Mintoff in an unpublished manuscript entitled *Gauthier on Intention and Action,* at 15. Gauthier suggests his acceptance of the objection in another manuscript, entitled *Odysseus and the Tortoise.*

rationally continue to issue and to execute deterrent threats, provided that maintaining the policy is indeed optimal. But before the policy has produced benefits, it cannot be rational to institute it, for one could end up committing oneself to a series of actions that might not turn out to be beneficial at all. Aware of this possibility from the outset, a rational agent can never justify adopting a policy that involves issuing and making good on deterrent threats.[20] And this means that when she must choose between performing a suboptimal act in furtherance of a policy and abandoning the policy altogether, the latter will seem most rational.

Although this brief summary of the difficulty with the appeal to policy is surely too abbreviated to be fully convincing, I wish to accept it as correct in the present context. For I wish to explore instead the supposed asymmetry between threats and assurances. If it turns out that there is no relevant asymmetry, it would not be *necessary* to try to make the appeal to policy successful anyway. So let us see if we can articulate more clearly the basis for thinking there is an asymmetry between assurances and threats.

What is the difference between assurances and threats, then, that seems to preclude appealing to the optimality of plans involving threats but not to plans involving assurances? One difference is that threats introduce an element of *risk* or *uncertainty*. When Clarence assures Doreen that he will help her plow her field next week if Doreen plows Clarence's field now, Clarence's future performance is not implicated unless Doreen actually assists. This means that there is no risk to Clarence that following through on the commitment to help Doreen will turn out to be disadvantageous, relative to the baseline of what things would have been like for Clarence had he not made the commitment. Where deterrent threats are concerned, however, the advantages afforded by issuing the threat are not certain to accrue. Where the threat fails, it is precisely because these advantages have not accrued that the threat must be carried out. The rationality of committing to a course of conduct involving threat-issuance and threat-execution thus depends on the threatener's assessment of the probability that she will not have to make good on the threat. The higher the risk of follow-through, the less rational issuing the threat becomes.

This makes the issuance of a threat look like a lottery, where the cost of the lottery ticket is the *ex ante* chance that the agent will have to make good on the deterrent threat multiplied by the cost of doing so, and the lottery prize is the successful deterrence of the person threatened. Whether it is rational to enter such a lottery should be determined by balancing the value of successful deterrence, discounted by the chances of winning that value, against the cost to the agent of having to make good on the failed threat, discounted by the chances of having to make good on it. Of course, if one is less than fully certain that the action one seeks to deter will be performed

20. I pursue another objection to the appeal to policy in *Rational Temptation* in PREFERENCES, PRINCIPLES AND PRACTICES: ESSAYS IN HONOR OF DAVID GAUTHIER (Chris Morris & Arthur Ripstein, eds., forthcoming 2000).

even *without* one's intervention, one should also take into account the likelihood of the action's being performed without intervention, as well as the costs of this occurring. A rational deterrent threat, then, is one in which the expected benefit of issuing the threat exceeds the expected costs associated with issuing it, assuming that a failed threat entails follow-through.

Prior to his attempt to account for the rationality of deterrent threats in terms of policies, Gauthier presented an account built roughly around this intuition, which he subsequently abandoned in favor of the appeal to policy. Gauthier's early account will prove more useful for our purposes than his later account, and let us therefore consider it in greater detail. Assume I am considering issuing a threat to do something, x, in order to deter you from doing something else, y. My options are to threaten to do x, and thus commit to doing x should the threat fail to deter you from doing y, or do something else, x'. You can either perform the action I wish to deter you from performing, namely y, or you can do something else, y'. Then, Gauthier's early account says, the following condition would have to be satisfied to make it rational for me to issue a sincere threat to do x:

$$[P_x u(yx) + (1 - P_x)u(y')] - [P_{x'}u(yx') + (1 - P_{x'})u(y')] > 0.$$

$P_x u(yx)$ is the probability, given that I threaten x, that you will do the undesired action y anyway, multiplied by the disutility to me of your doing it. $(1-P_x)u(y')$ is the probability, given again that I threaten x, that you are successfully deterred from doing y, multiplied by the utility to me of your being deterred. $P_{x'}u(yx')$ is the probability, given that I choose *not* to threaten x, of your performing the undesired action y, multiplied by the (dis)utility to me of your doing so. Finally $(1 - P_{x'})u(y')$ is the probability, given again that I choose not to threaten x, of your choosing to perform some action other than y, namely y', anyway. The first expression in square brackets $[P_x u(yx) + (1 - P_x)u(y')]$ represents my expected utility should I sincerely threaten to do x if you do y. The second expression in square brackets $[P_{x'}u(yx') + (1 - P_{x'})u(y')]$ represents my expected utility from embarking on some other course of action instead. The entire inequality indicates that it is rational to issue a deterrent threat when the expected benefits from issuing the threat are greater than the expected costs, or, as written, when the benefits minus the costs are greater than zero.[21]

21. David Gauthier, *Deterrence, Maximization, and Rationality, in* THE SECURITY GAMBLE 100 (1984). Gauthier makes the point by saying that it is rational to issue a sincere deterrent threat when the proportionate decrease that the deterrent policy effects in the probability of facing the other's undesired action is greater than the minimum required probability for deterrent success. He represents this as follows:

$$[(P_{x'} - P_x)/P_{x'}] > P.$$

$(P_{x'} - P_x)/P_{x'}$ represents the increased chances that the other party will refrain from performing the undesired act, given that one threatens him, divided by the absolute chances of deterring him. The P on the right-hand side of the equation is the point at which the agent is indifferent between performing the deterrent action and not performing *ex ante*.

Gauthier's early account proposes an *ex ante* measure of when it is rational to issue a sincere deterrent threat: It is rational to issue a threat when the expected value from issuing it exceeds its expected costs. Let us call this the *ex ante* solution to the rationality of deterrent threats. The *ex ante* solution would clearly be the rational way to assess the merits of issuing deterrent threats if one could precommit to following through on the threat. For in that case the decision to threaten and the decision to follow-through are folded into a single, up-front choice, and the payoffs associated with that choice are exhausted by the expected utility from the benefits and costs associated with issuing the deterrent threat. The trouble is that the account will not serve to vindicate the rationality of deterrent threats when the execution of the threat must take place *deliberatively,* that is, pursuant to a decision to execute a failed threat. For a threat may be justified by its expected value *ex ante,* even when carrying out the threat *ex post* cannot be justified in this way. Reflecting at the point of execution on the benefits to me of the threat and its associated follow-through, I cannot regard the *ex ante* benefits of issuing the threat as supplying any reason to execute the threat once those advantages have failed to accrue. The problem with Gauthier's early account, then, is that it operates entirely *ex ante,* and *ex ante* reasoning will only justify automatic threat-execution. It cannot justify executing threats in a way that is sensitive to ongoing deliberation, particularly where the execution of the threat must be decided on *after* the hoped-for benefits from issuing the threat have failed to accrue.

How to accommodate deliberative threat-execution will become clear if we recall how deliberative assurance-execution was accommodated in the case of an assurance. To make the case of an assurance parallel to that of a deterrent threat, let us consider the rationality of an assurance that also contains an element of risk. That is, let us re-create the threats problem on the assurance side by making the benefits from the assurance less than certain to accrue. Suppose, for example, that Clarence's field is twice as large as Doreen's field. Doreen is not willing, therefore, to offer her assistance to Clarence in exchange for assistance plowing her own field. She would, however, be willing to offer Clarence a 50% chance of assistance if Clarence would commit to helping Doreen plow her field next week. Both parties should regard this as a fair deal.

Next, suppose Clarence flips a coin to determine whether he will receive assistance, and he loses the gamble. Can he *now* apply the deliberative principle from the original assurance case to vindicate the rationality of making good on the commitment to assist Doreen? That principle, to recall, said that an agent should perform those actions demanded by plans it was optimal to adopt. In this case, the plan of exchanging a fair shot at Doreen's assistance for his own labors was the optimal plan for Clarence to adopt. True, he now wishes he had never exchanged a commitment to assist Doreen for a fair shot at her assistance. But there is nevertheless a sense in which he has received exactly what he bargained for when he entered the agreement, namely a 50%

chance of receiving Doreen's assistance. If the original deliberative argument for making good on assurances is correct, then the same argument should apply here: Act consistently with a prior optimal plan if the circumstances are among those envisioned at the time the plan was adopted.

Consider, alternatively, the famous example of the eccentric billionaire who promises to put a million dollars in your bank account by midnight tonight if you form the intention to drink a toxin tomorrow. The toxin will leave you sick for 24 hours but will have no long-lasting effects.[22] The billionaire has a nearly infallible way of telling whether you have formed the requisite intention. The rational course is arguably to commit to drink the toxin and then actually to drink it, because you are better off under the plan that involves intending to drink and drinking than you would be without the plan. Once again, the relevant deliberative principle says it is rational to optimize over plans and to select particular actions by restricting your reasoning to the plan it was optimal to adopt.[23]

Now consider a variant on this puzzle. Suppose the billionaire presents you with a gamble. He will use a randomizing device that will give you a one-in-a-hundred chance of having a million dollars deposited in your bank account tonight should you form the intention by midnight to drink the toxin tomorrow. The toxin is milder than before: It will leave you with an unpleasant headache for several hours but will have no other untoward effects. You regard the trade-off of headache for the one-in-a-hundred chance at a million dollars as worthwhile, and so you now commit to drink the toxin tomorrow. Suppose you learn at midnight that you have lost the gamble and that the money is not in your bank account. Arguably *that* should not make the difference to whether it is rational for you to drink the toxin tomorrow. For if it was irrelevant to the rationality of your drinking in the original case that nothing you could do after midnight would affect whether the money was or was not in your bank account, then it should be equally irrelevant in the modified case. If *anything* made it rational for you to drink the toxin in the original case, after all, it was not that you would be glad immediately prior to drinking that you had adopted the plan of committing to drink and drinking. Rather, it is that the plan of committing yourself to drinking and then actually drinking was the optimal plan to adopt. This fact about the plan—its optimality—remains true even if it turns out that the consequences of having adopted the plan are not the ones for which you had ideally hoped, assuming once again that the circumstances are among those you envisioned when you adopted the plan initially.

It looks, then, as though injecting an element of risk into the gains from assurances should not fundamentally affect the rationality of making good on them. Following through on an assurance, whether risky or not, is ra-

22. The problem was first posed by Gregory Kavka, *The Toxin Puzzle,* 43 ANALYSIS 33 (1983).

23. *See* David Gauthier, *Resolute Choice and Rational Deliberation: A Critique and a Defense,* 31 NOÛS 20 (1997).

tional if the plan under which the assurance was issued is the plan it was optimal to adopt. The argument easily generalizes to threats, for there is no relevant difference between a threat and a risky assurance. In Section VI I will formalize this understanding of threats and risky assurances and show more precisely how it accommodates deliberative threat-execution. What we require first, however, is an argument for the suggestion that there are gains in the case of a failed risky assurance or a failed threat that are comparable to the gains from a sure assurance. The question, in other words, is whether an *ex ante* increased chance of benefit is *in and of itself* a benefit.

Before turning to this question, it is worth noticing that the foregoing may ironically provide ammunition for someone who wishes to reject Gauthier's original claim about the rationality of making good on assurances. For if we must accept following through on assurance "gambles" as rational if we accept following through on sure assurances as rational, then for someone who thinks the rationality of the former cannot be vindicated when assurances must be *deliberatively* executed, the rationality of suboptimal actions in fulfillment of assurances as part of optimal plans cannot be defended after all. But in this article I am assuming, rather than arguing for, Gauthier's original claim about assurances. This is not because I am certain it is correct, but rather because I wish to consider the light it sheds on the rationality, and ultimately on the morality, of deterrent threats. I will not, therefore, attempt to address this possible implication of my argument in the present context.

V. IS AN *EX ANTE* INCREASED CHANCE OF BENEFIT ITSELF A BENEFIT?

Should you think of yourself as benefitted if I make you a gift of a lottery ticket? One answer is that whether I have benefitted you depends on whether I have bought you the winning ticket. For surely the ticket is no benefit to you, considered in itself. An alternative view of the situation, however, seems no less plausible. You might regard my giving you the ticket as a benefit whether or not I have bought you the winning ticket. For the ticket increases your chance of benefit, and that in itself may be a benefit.

The same problem arises where *harm* is concerned. Is exposing someone to an increased risk of harm itself a harm? There is some support for the suggestion that it is. Although in law, courts and commentators have mostly rejected the claim that an increased risk of harm is compensable,[24] some interesting cases have gone the other way. In *Hymowitz v. Eli Lilly & Co.*,[25] for example, plaintiffs were compensated for damage allegedly sustained owing to their mothers' ingestion of DES during pregnancy. The court allowed recovery according to the *amount of risk* each defendant created to

24. *See, e.g.*, Plummer v. Abbott Laboratories, 568 F. Supp. 920, at 922 (D.R.I. 1983); *see also* Ernest Weinrib, *Causation and Wrongdoing*, 63 CHI.-KENT L. REV. 407 (1987).

25. 539 N.E.2d 1069 (N.Y. 1989).

the public from DES. The correct measure of this risk, the court said, can be established by the principle of market-share liability introduced in *Sindell.*[26] But the court also suggested that because it was interested only in the risk of harm, market share should not be determined along causal lines, as it had been in *Sindell.* Consequently, the court refused to exculpate defendants who were part of the market producing DES but who could prove that they could not have *caused* plaintiff's injuries. As the court said:

> [B]ecause liability here is based on the over-all risk produced, and not causation in a single case, there should be no exculpation of a defendant who, although a member of the market producing DES for pregnancy use, appears not to have caused a particular plaintiff's injury.[27]

In other words, the court used market-share liability as a justification for compensating for exposure to risk, rather than as a way of estimating each defendant's contribution to resultant (i.e., non-risk-based) harm.

There are also a number of decisions allowing recovery for emotional distress where the distress was caused by the defendant's imposition of an unjustified risk of harm.[28] Courts stress in such cases that there would be no recovery for emotional distress in the absence of exposure to risk. And although there is often a requirement that the plaintiff manifest symptoms of emotional suffering, this may once again be merely a way of measuring damages from exposure to risk. Arguably, the harm for which plaintiffs are compensated is the exposure itself, rather than the emotional distress that resulted.

Among commentators, the idea of compensation for risk has gained some support in recent years. One author, for example, argues that failure to compensate for risk in cases in which the exposure to risk is known many years before the risk could possibly eventuate imposes a substantial hardship on plaintiffs.[29] Another insists on a related point, namely that the duty to pay damages should be tied to risk-creation rather than to the infliction of injury.[30] The duty to compensate should be tied to the offender's wrong-

26. Sindell v. Abbott Labs., 607 P.2d 924 (Cal. 1980) (assigning liability to drug manufacturers according to their share of market at time of plaintiff exposure to product, rather than requiring plaintiff to demonstrate defendant's actual causal contribution).

27. *Hymowitz,* 539 N.E.2d at 1078.

28. *See, e.g.,* Haggerty v. L&L Marine Services, Inc., 788 F.2d 315, 317–18 (5th Cir. 1986); Dunn v. Owens Corning Fiberglass, 774 F. Supp. 929, 942 (D.V.I. 1991).

29. Glen O. Robinson, *Probabilistic Causation and Compensation for Tortious Risk,* 14 J. LEGAL STUD. 779, 785–86 (1985).

30. Christopher H. Schroeder, *Corrective Justice and Liability for Increasing Risks,* 37 U.C.L.A. L. REV. 439 (1990) ("The *ex ante* rationale for not waiting for actual harm to occur is that the moral quality of the agent's behavior turns on nothing that subsequent events can reveal.") Many commentators reject the claim. Stephen Perry, for example, argues convincingly that the idea that risk is a harm suggests a commitment to an objective account of risk. He also suggests, however, that the objective account of risk is less plausible than a subjective account, and accordingly suggests that risk should not be considered a harm in and of itself. Stephen Perry, *Risk, Harm and Responsibility, in* PHILOSOPHICAL FOUNDATIONS OF TORT LAW 321 (David Owen ed., 1995).

doing, and someone who imposes a risk of harm has behaved equally wrongfully whether or not his or her conduct results in actual injury. Granted, an advocate of this way of looking at the duty to compensate in tort is not, strictly speaking, required to think that to expose someone to risk is to harm him. But a natural way to understand this suggestion is that a person acts wrongly when subjecting another person to risk, among other things, because to expose someone to risk is to harm him.[31]

The analogous point on the benefit side also receives some support. There are, for example, "lost-chance" cases, in which courts have allowed plaintiffs to recover for the loss of a chance of future benefit.[32] Courts maintain that damages in such cases are a function of the total amount of damage from injury or death, multiplied by the percentage of the chance lost.[33] The lost-chance cases thus abandon the traditional causation requirement that a plaintiff's injuries were more likely than not to have resulted from the defendant's negligence. Instead, the plaintiff is compensated for the chance that the defendant's behavior deprived her of a benefit.

For our purposes, it is not necessary to accept the legal trend toward compensating for exposure to risk or for lost chance of benefit. For one need not think defendants should have to *compensate* for exposure to risk in order to accept the claim that an *ex ante* increased risk of harm is itself a harm. Nor need one think that defendants should have to compensate for depriving a plaintiff of a lost chance of benefit in order to think of an increased chance of benefit as itself a benefit. No one, after all, thinks that all harms, still less all lost benefits, should be compensable. But the legal trend toward compensating for risk or for lost chance of benefit does lend support to the idea that exposure to risk is a harm, and similarly that an increased chance of benefit is itself a benefit. Indeed, the law recognizes the idea where compensation is not at issue as well. Statutes criminalizing reckless driving irrespective of outcome, for example, reflect the intuition. And we may accept it in ordinary morality when we believe it appropriate to be angry with someone who has exposed us to an unjustified risk of harm, or who has needlessly deprived us of a chance of benefit, whether or not actual damage resulted from the exposure or whether the actual benefit was lost.

31. As Ken Simons points out, Schroeder's position on this point is curious, for he thinks the duty to pay should be triggered by risk-creation, but he does not think that plaintiffs should recover unless they have suffered resultant injury. *Id.* at 467–68. *See* Kenneth Simons, *Corrective Justice and Liability for Risk-Creation: A Comment*, 38 U.C.L.A. L. REV. 113, 120–25 (1990). In a response to Simons, Schroeder raises the question of whether risk of harm is itself a harm, but declines to answer it. Christopher Schroeder, *Corrective Justice, Liability for Risks, and Tort Law*, 38 U.C.L.A. L. REV. 143, 160 (1990).

32. *See* Roberts v. Ohio Permanente Med. Group, Inc., 668 N.E.2d 480 (1996) (damage from lost chance of recovery allowed to reach jury if plaintiff can show lost chance resulted from defendant's negligence, even where chance of recovery would have been substantially less than 50% in the absence of defendant's negligence).

33. *Id.* at 480.

VI. THE *EX POST* ACCOUNT

If we think of an *ex ante* increased chance of benefit as itself a benefit, threats can be assimilated to sure assurances. For the agent who must execute a failed threat can say to herself that the benefit she was seeking by issuing the threat was the increased *chance* of deterrence, and *that* she has received. At the point of execution, the agent should apply precisely the deliberative principle we saw with assurances: She should weigh the benefit of the *ex ante* increased chance of deterrence against the cost of making good on the failed threat, and if the balance of benefit over cost is favorable, the plan she adopted is optimal, and she should proceed to act as it requires. Recall that in the case of a sure assurance, we measured the benefit already gained—the advantage Clarence received from having Doreen's help plowing his field—against the cost of future performance—the inconvenience of helping Doreen plow her field. So in the case of a failed threat, we should weigh the benefit already gained—the *ex ante* increased chance of deterring *B*—against the cost of future performance—the cost of making good on the threat to tell Burt's spouse of the affair. As in the case of sure assurances, the threats that are rational to execute, and hence the threats that are rational to issue, are those for which the *ex ante* benefits outweigh the costs of threat-execution.

On this way of looking at the matter, we can express the rationality of issuing a deterrent threat as follows:

$$[(1 - P_x)u(y') - (1 - P_{x'})u(y')] > C.$$

$(1 - P_x)u(y')$ is the probability, given that I issue a threat to x, that you will be deterred from doing the undesired action, y. It is thus the expected benefit from issuing the deterrent threat. $(1 - P_{x'})u(y')$ is the probability that you will not perform the undesired action, y, given that I choose to do something other than threaten x. It is thus my expected benefit from adopting any course of action other than threatening x, namely x'. C is the absolute cost of making good on the deterrent threat should it fail to deter. The left-hand side of the inequality represents the expected benefit to the agent of issuing a deterrent threat, and the right-hand side represents the cost of having to carry through with the threat should it fail to deter. The cost, in turn, can be expressed as $u(yx') - u(yx)$, which is the (dis)utility of foregoing another course of action other than making good on the deterrent threat, minus the (dis)utility of having to make good on the deterrent threat by performing x. The left-hand side of the inequality reduces to:

$$(P_{x'} - P_x)u(y'),$$

and so the entire inequality is:

$$(P_{x'} - P_x)u(y') > u(yx') - u(yx).$$

Compare this result with the *ex ante* solution. Recall that the problem we noted with that account was that it treats threat-execution as though it were automatic. It thus fails to accommodate the two-tier deliberative principle Gauthier advances in the case of assurances. Instead of determining when it is rational to execute a deterrent threat by considering when it is rational to issue such a threat in an up-front, *ex ante* decision, the account proposed here determines the rationality of issuing a deterrent threat by considering when the plan of issuing and if need be executing a deterrent threat would be rational to adopt, and then uses this as a criterion for determining when it would be rational to make good on such a threat *ex post*. Just as in the case of sure assurances, it would arguably be rational to make good on a deterrent threat when the expected benefits of having issued such a threat exceed the absolute cost of having to make good on it. For want of a better name, let us call this the *ex post* account of deterrent threats, by contrast with Gauthier's *ex ante* account.

Recall that the *ex ante* account was expressed as:

$$[P_x(yx) + (1 - P_x)u(y')] - [P_{x'}u(yx') + (1 - P_{x'})u(y')] > 0.$$

We can rewrite Gauthier's inequality by leaving only terms that contain $u(y')$ on the left and moving all other terms to the right:

$$[(1 - P_x)u(y') - (1 - P_{x'})u(y')] > [P_{x'}u(yx') - P_xu(yx)],$$

and then simplifying the left-hand side to produce:

$$[(P_{x'} - P_x)u(y')] > [P_{x'}u(yx') - P_xu(yx)].$$

Compare this to our inequality, which was:

$$[(P_{x'} - P_x)u(y')] > [u(yx') - u(yx)],$$

and the difference between the *ex ante* and the *ex post* accounts should be clear. The *ex post* account is exactly the same as the *ex ante* account on the left-hand side of the inequality, which is a measure of the benefit of the deterrent threat *ex ante*. The solutions, however, differ on the cost side. The *ex ante* solution discounts the cost by the probability that it will not turn out to be necessary to execute the deterrent threat. The *ex post* solution, by contrast, does not discount the costs, and thus it measures the costs associated with issuing deterrent threats from the standpoint of someone deciding what to do once his threat has already failed. In this way, the *ex post* account accommodates deliberative threat-execution, as it provides an answer to the question of when threat-execution itself is deliberatively rational, and recommends the issuance of only those threats that satisfy this condition.

What are the consequences of this difference? The *ex ante* account, which discounts both benefit and cost, would serve to vindicate the rationality of many more deterrent threats than would the *ex post* account, because the right-hand side of the inequality will always be larger in the *ex post* account than it is in the *ex ante* account.[34] The *ex post* account would vindicate the rationality of only a small number of deterrent threats. This is as one might expect, given that the *ex post* account does not allow us to justify threat-issuance and threat-execution in a single, up-front choice. It justifies fewer threats because it imposes an additional condition, namely that threats must be deliberatively executed.

We can see the difference between the *ex ante* and the *ex post* accounts if we return to one of our earlier examples and assign some values to the various options. Recall that Abigail wants to deter Burt from applying for a job that Abigail hopes to obtain, and thus threatens to tell Burt's spouse that Burt is having an affair. Let us suppose that deterring Burt from applying for the job is worth roughly $1,000 to Abigail (that is, she would pay that sum of money to be sure of deterring Burt from applying). And suppose it will cost Abigail roughly $300 if she must make good on the threat to tell Burt's wife of his affair (say, she would pay that sum of money to avoid having to disclose the affair). Suppose further that there is a 0.8 chance that Burt will apply for the job if Abigail does not issue the threat (or a 0.2 chance that he will fail to apply for it if left to his own devices), and a 0.4 chance that he will be dissuaded from applying for the job if Abigail does issue the threat (or a 0.6 chance that he will still apply for the job, even in the face of the threat). Then the *ex ante* account would suggest it is rational to issue the threat if the following inequality holds:

$$[(.8 - .6)1{,}000] > [.8(0) - .6(-300)].$$

So we have:

$$200 > 180.$$

The benefits of issuing the threat outweigh its costs from an *ex ante* perspective, and the threat is rational to issue. On the *ex post* account, the left-hand side of the inequality will stay the same, but the right-hand side will not be discounted by the probabilities. The *ex post* account would accordingly suggest that it is rational to issue the threat if the following inequality holds:

$$200 > 300,$$

which it does not. Thus, on the *ex post* account the threat would not be rational to issue.

34. This can be shown using the fact that $u(yx')$ is a cost or disutility, and so should be represented by a number less than or equal to zero.

The fact that many fewer threats will be rational to issue on the *ex post* than on the *ex ante* account is an advantage of the former account, as it tracks our intuition that it is rational to follow through on, and hence to issue, only those threats where the benefits *far* outweigh the costs. For example, it would provide an explanation of the sense many have that issuing credible nuclear retaliatory threats is not rational: The costs on the right-hand side are so high that the expected gains from issuing such a threat are unlikely to be sufficiently great to make the course of conduct worthwhile.

The *ex post* solution applies the same reasoning to determine the rationality of issuing assurances, including risky assurances, that it uses for threats. The difference between a risky assurance and a threat, on the one hand, versus a sure assurance, on the other, would simply be that the left-hand side of the inequality would not discount by probabilities where sure assurances are concerned—that is, the benefit received would not have to be reduced by the chances of its nonoccurrence. But this does not affect the structure of the account: An assurance, whether risky or sure, is rational to issue if the benefits from issuing it are greater than the costs associated with making good on the assurance. When this condition is satisfied, it is rational to make good on an assurance rationally issued, whether the benefit received was only an increased chance of benefit or an actual, resultant benefit.

The asymmetry between threats and sure assurances, then, is not a deep one, and it does not justify the divergent solutions Gauthier proposed in his article *Assure and Threaten.* The appearance of asymmetry between the two stems from the fact that, as in the case of risky assurances, the benefit side of the inequality where threats are concerned is a discounted measure. This suggests that it will be harder to vindicate the rationality of threat-execution, and consequently harder to vindicate the rationality of issuing a deterrent threat, than it will be to vindicate the rationality of a comparable sure assurance. But this is only a matter of application. It does not reflect an asymmetry at the level of the basic structure of the rationality of assurances and threats.

VII. JUSTIFICATION REVISITED

We are finally in a position to return to the question of the justifiability of deterrent threats. In particular, I wish to consider what an *ex post* account of the justifiability of deterrent threats would look like, developing the parallel between the rationality and the morality of such threats. The moral equivalent of the *ex post* account would hold that a deterrent threat is justifiable when the expected moral benefits of issuing the threat are sufficient to outweigh the absolute moral cost of having to make good on the threat should it fail to deter. Thus, the justifiability of a police officer's use of deadly force against a fleeing felon would depend on whether the benefit

of deterring the suspect, discounted by the *ex ante* chance that he will fail to be deterred, is a sufficient moral benefit to overcome the independent wrongfulness of using deadly force against the suspect. Just as the rationality of deterrent threats was determined in the *ex post* account according to the balance of costs and benefits considered from the standpoint of someone who must make good on a threat already issued, so the morality of deterrent threats should be determined by the balance of moral costs and benefits determined from the *ex post* standpoint as well. In weighing moral costs and benefits, we should not discount the moral costs of having to make good on a threat by the probability that it will not be necessary to execute the threat after all. The act of violence performed in execution of the threat must itself be morally justifiable, not on independent moral grounds, but as a "price" paid for the *ex ante* moral benefit of potentially deterring the wrongful conduct in question.

Several features of the police practice admit of ready explanation under an *ex post* moral account. First, the more dangerous the suspect is believed to be, the more beneficial a given chance of deterring him would be, and thus the more justified the issuance of the deterrent threat against him. The modern constitutional requirement of dangerousness might be thought of as a bright-line rule that captures this intuition. According to the account provided here, however, for the use of deadly force to be justified against suspects on preventive grounds, the chances of deterring them from flight would have to be significant, and any such threat would be extremely difficult to justify if the officer thought it likely that the use of force under the circumstances was likely to result in death.

Second, we would have at least a partial explanation for the sense that putting a suspect on notice of the impending harm helps to justify it. Notice is essential for establishing the increased *ex ante* chance of deterrence from the issuance of the threat. On an *ex post* account, the chances of deterring the wrongful act must be significant for a deterrent threat to be justified. The more effective the threat the more easily justified it is, for the expected benefits would increase with the likelihood of deterrent success. There is an irony to this feature of the account, for presumably a person can increase the chances of deterrent success simply by increasing the severity of the harm threatened. And this makes it look as though the more severe the harm threatened, the more justified the threat.[35] But recall that the severity of the threatened harm will also figure in the cost side of the account, and indeed will figure there more heavily than it will on the benefit side, because it will not be discounted by the probability of deterrent success. Hence, the easiest threat to justify will be one that involves little harm should it fail to deter but whose chances of deterrent success are nevertheless high. This seems an intuitively plausible result.

35. I am grateful to Daniel Rodriguez for this suggestion.

Finally, the account I have offered suggests more palatable conclusions about a range of possible practices than the Extreme Account was able to provide. For the proposed account places limits on both the rationality and the morality of deterrent threats in a way that the Extreme Account did not: The act must follow from a plan in which the balance of moral benefit over moral burden is favorable. This requirement limits the acts it is possible to justify by bootstrapping ourselves up from the justifiability of issuing deterrent threats. We thus have a basis for distinguishing the police practice of threatening and sometimes using deadly force against fleeing felons from threatening to machine-gun trespassers on the courthouse lawn, along with actually machine-gunning them should the deterrent threat fail. In the latter case, it is plausible to suppose that although the *ex ante* chances of deterring the wrongful trespassing may be relatively high, the overall *ex ante* moral benefit is not great. And this is because the wrongful act whose deterrence is sought—relatively harmless trespassing—is not sufficiently damaging to make deterrence a great moral benefit. Conversely, the harm required should the deterrent threat fail is grave: Using deadly force against individuals who pose little likelihood of harm to society is a great moral wrong, one whose infliction could only be justified in the face of a wrongful act of substantial magnitude.

What the solution to the rationality of deterrent threats provides, by way of analogy, is not a complete answer to the question of when practices involving the use of deterrent threats are justifiable, but rather a framework for thinking about practices in which we cannot offer an independent justification for the acts of violence they involve. This framework, however, leaves much unresolved, perhaps so much that it will seem of little value, since it is unable to provide the kind of practical guidance that, for example, the parallel account of the rationality of deterrent threats is able to provide. While it is clear how to determine costs and benefits where the rationality of deterrent threats is concerned, quantifying *moral* costs and benefits seems a hopeless task.

I cannot hope to solve this problem, for unlike where rational payoffs are concerned, moral "payoffs" could never admit of any great precision. For this reason, an *ex ante* moral account can only provide a rough and ready way of thinking about the moral aspects of a course of action involving threats. But even if taken in this spirit, we still should be able to say something about what would *count* as a moral cost or benefit. This will depend on what moral theory we use to assess plans involving the issuance of deterrent threats. Let me then in closing return to the outstanding question of what theory of morality could do the work in vindicating the morality of deterrent threats that the pragmatic theory of rationality did in vindicating their rationality.

One idea would exploit the parallel between the rational and the moral dimensions of deterrent threats. The existence of the parallel may suggest that the correct moral theory by which to evaluate the moral costs and

benefits is itself a theory that has rationality at its core. I wish to illustrate the idea I have in mind by considering briefly a further example of a preemptive practice from outside the criminal law: the use of penalty clauses in contracts.

Suppose you and I are negotiating a contract for you to paint my living room walls. And suppose it is not terribly important to me to paint my living room now, as opposed to next year, but it is terribly important to me *if* I undertake the project now that it be completed by next Thursday. You are a painter who is planning on going out of business next year, and it is important to you to obtain the contract for work *now.* It is not important to you, however, that the project be finished by next Thursday. To induce me to paint my living room now, you offer me a contract in which you will paint my living room for $300 by next Thursday, with a promise to pay me $300 for every day's delay thereafter.

A court might refuse to enforce the delay provision if it suspects it constitutes a penalty for breach rather than a liquidated damages clause.[36] The explanation for the rule against penalty clauses is that the harm they inflict on the party in breach is disproportionate to the losses the victim of breach sustains. Courts, in short, are bothered by the fact that there is no independent justification for forcing the party in breach to make a payment so in excess of actual damages. But the ability to include a threat of substantial harm in a contract may be highly advantageous for both parties, for it may make possible a mutually beneficial agreement that would not be available in the absence of such a threat. It may be advantageous to both of us to enter into the contract to paint my living room, but this mutual advantage can only be realized if I can receive adequate assurance that you do not intend to delay. In the context of a freely negotiated contract, the fact that it would be to the advantage of both parties, and hence *rational* for both parties to incorporate such a clause, should provide an adequate justification for enforcing such clauses, as long as no third party is thereby harmed. If both parties to the contract would be better off, and no one would be worse off, there is no reason that threats to induce future performance should not be enforceable.[37]

One possible way of justifying preemptive practices would parallel the above justification for allowing the use of penalty clauses in contracts. The harm inflicted in each case cannot be justified on independent grounds. But it may be justified on the grounds that there are advantages to allowing agents to make use of deterrent threats in their dealings with one another. Perhaps, then, preemptive practices can be justified as penalty clauses writ large: They are morally defensible because they provide gains that those affected by them can regard *ex ante* as beneficial given their individual

36. E. Allan Farnsworth, FARNSWORTH IN CONTRACTS § 12.18 (1990).

37. Some contract scholars support the use of penalty clauses within certain bounds. *See* Melvin Eisenberg, *The Limits of Cognition and the Limits of Contract,* 47 STAN. L. REV. 211 (1995).

concerns. A practice involving the issuance and execution of deterrent threats may thus be mutually beneficial for the members of a society who must be governed by it.

We cannot think of the fleeing felon rule as the product of a rational agreement for mutual benefit between police officer and fleeing suspect. But considering the matter more broadly, it seems plausible to suppose that members of a society interested in protecting individual liberty while providing themselves with security from wrongful violence would regard certain rules for evaluating the culpability of those suspected of such violence as advantageous. In particular, rational agents might regard themselves as benefitted by a rule for the arrest of subjects that leaves the decision of whether to be exposed to violence in a certain sense up to the suspect: Because the officer must issue a threat prior to using force, the suspect is on notice of the consequences of flight. He can therefore choose to avoid those consequences if he wishes by conforming to the officer's order.

This suggests that the correct moral principle for the evaluation of the preemptive practices may be a principle of agreement, rather than some independent principle of justice. One must treat this conclusion gingerly, however. For the peculiar logic of the preemptive practices does not appear to generalize to other institutions, institutions that for the most part are guided by such independent principles. Independent *preventive* principles, for example, appear to govern a practice like self-defense. And independent *retributive* principles seem most essential to our punitive practices. Indeed, the anomalous nature of the preemptive practices might give us pause: Is mutual advantage sufficient justification for a practice involving acts of violence that cannot be independently morally justified after all?