

University of Pennsylvania Carey Law School

Penn Law: Legal Scholarship Repository

Faculty Scholarship at Penn Law

2013

Pragmatic Rationality and Risk

Claire Oakes Finkelstein

University of Pennsylvania Carey Law School

Follow this and additional works at: https://scholarship.law.upenn.edu/faculty_scholarship



Part of the [Cognition and Perception Commons](#), [Law Commons](#), and the [Philosophy of Mind Commons](#)

Repository Citation

Finkelstein, Claire Oakes, "Pragmatic Rationality and Risk" (2013). *Faculty Scholarship at Penn Law*. 1538. https://scholarship.law.upenn.edu/faculty_scholarship/1538

This Article is brought to you for free and open access by Penn Law: Legal Scholarship Repository. It has been accepted for inclusion in Faculty Scholarship at Penn Law by an authorized administrator of Penn Law: Legal Scholarship Repository. For more information, please contact PennlawIR@law.upenn.edu.

Pragmatic Rationality and Risk*

Claire Finkelstein

Pragmatic theories focus on whether agents fare better acting on the basis of a particular intention or plan, rather than whether this can be justified in terms of the expected utility associated with the plan. This article argues that, while attractive, pragmatic theories have difficulty vindicating the rationality of plans involving an element of risk. In "Assure and Threaten," David Gauthier noticed this difficulty with respect to deterrent threats. This article argues that the same difficulty exists for assurances involving an element of risk. It then explores whether Pragmatists could solve the shortcomings of their approach by adopting the Chance Benefit Thesis, namely, the thesis that a chance of benefit is itself a benefit.

I. OPTIMAL PLANS AND THEIR SUBOPTIMAL SUBPARTS

Much recent work on practical rationality has focused on intentions and actions in the lives of rational creatures. A common starting point for discussion is the observation that intentions have many of the characteristics of plans. More precisely, intentions appear to be partial plans, in that they require the formation of other, more immediate intentions for their fulfillment.¹ By forming intentions and planning for the future in light of those intentions, a rational agent reasons in stages: she first settles on a plan that is optimal in light of her beliefs and desires and then restricts her reasoning to actions that will contribute to the execution of that prior plan. I shall call plans that are optimal in light of an agent's beliefs and desires "optimal plans." A practically rational agent, then, is someone who first restrains the intentions she forms to the confines of optimal plans and who then constrains her actions to those particular intentions.

* I wish to thank Michael Bratman, Geoff Brennan, Michael Finkelstein, David Gauthier, Joe Mintoff, Christopher Morris, Connie Rosati, Seana Shiffrin, and Bruno Verbeek for their helpful comments at various stages in the drafting of this article and also audiences at York University, the University of Newcastle (Australia), and the University of Pennsylvania philosophy department.

1. Michael Bratman has most compellingly articulated the role of plans in practical reasoning. *Intentions, Plans and Practical Reason* (Cambridge, MA: Harvard University Press, 1987).

Ethics 123 (July 2013): 673–699

© 2013 by The University of Chicago. All rights reserved. 0014-1704/2013/12304-0007\$10.00

There is a complication, however. Sometimes the optimal plan will require an agent to perform actions that are not, in and of themselves, optimal. Indeed, sometimes the optimal plan requires actions that are distinctly suboptimal in execution. The need to perform individual suboptimal actions, however, does not threaten the optimality of the plan. As long as the agent is better off under the plan than had he never adopted the plan in the first place, the plan remains optimal.

In the usual case, the actions I must perform to realize my plans have instrumental value: they are a means to realizing something else I value. But the relation between my plan and the actions required by my plan is not always instrumental. In some cases, the suboptimal action I must perform to realize my plan must be performed after the benefits of the plan have already accrued. The question such cases raise is whether it is rational for an agent to perform an action required by an optimal plan when the action is not instrumentally required to achieve the benefits of the plan.

Consider the case of the so-called Humean farmers. Alfred's field is ready for plowing this week, and Bertram's field will be ready next week. Neither farmer can plow his field by himself. And both would be better off helping the other and receiving help himself than he would be in the absence of any reciprocal arrangement between them. Alfred proposes that Bertram help him plow his field now, and in exchange Alfred will help Bertram next week. Is it rational for Bertram to accept Alfred's offer? That depends on whether it is rational for Alfred to keep his promise to help plow Bertram's field next week: if it is not rational for Alfred to plow Bertram's field, it is not rational for Bertram to render assistance first, given that this would likely result in Bertram's finding himself worse off than he would be in the absence of an agreement with Alfred in the first place. The rationality of the exchange thus depends on whether it is rational for Alfred to plow Bertram's field once Alfred has already secured Bertram's assistance with his own field.

In order to address the interesting problem cases of this sort raise, we must make the following assumptions: First, Alfred would achieve no reputational benefits from plowing Bertram's field, nor any other intangible benefit that would alter the payoffs from plowing. Second, there will be no future course of dealings between Alfred and Bertram that would make it advantageous for them to cooperate now. Third, there is no legal or other coercive enforcement of an agreement the two farmers might make. Any agreement between them must be adhered to on the basis of considerations of rationality endogenous to the plan. Fourth, common knowledge of rationality obtains between the parties, meaning that each individual is rational and each has knowledge of each other's rationality. The question, then, is whether it is rational for Alfred and

Bertram to enter into a sincere agreement to assist one another with plowing, and to maintain that agreement in the face of considerations that speak in favor of defection, pursuant to a mutual cooperation agreement the two might make.

As is well known, the standard economic approach to rationality maintains that cooperation is not rational under these circumstances. Once next week has arrived, Alfred will have nothing to gain and everything to lose by helping Bertram. While he does better under the bargain with Bertram than he would without it, there are no further gains to Alfred from abiding by the agreement after Bertram has performed. Alfred will not plow Bertram's field next week; therefore, since Bertram knows that Alfred is rational, Bertram will not agree to the exchange.²

In this paper, I shall explore some aspects of an alternative answer to the problem the farmers face and others of its ilk. David Gauthier, Edward McClennen, and others have advanced what they call the "pragmatic" theory of rationality, namely, a view that assesses the rationality of an act or plan by its ultimate utility to the agent.³ Proponents of the pragmatic theory accordingly hold that a plan that leaves an agent worse off than another available plan cannot be rational to adopt. It is rational for Alfred to plan to help Bertram plow his field next week, since under that plan, Alfred would be better off than in the absence of the plan, which would result in his plowing his field alone. The pragmatic account will thus diverge from standard expected utility theory in cases in which the plan that will make an agent's life go best requires him to perform acts that are not best, considered in isolation from a broader course of action. Since agents who optimize over plans rather than over individual acts can expect to fare better than agents who attempt to maximize act by act, the rational agent has pragmatic reasons to perform suboptimal actions in cases where required to carry out optimal plans. This may be so, even if those actions do not bear an instrumental relation to the benefits from the plan.

2. This is the standard economic result in this kind of situation, but a caveat is in order. The standard theory maintains that if there is repeat play, the parties should be able to cooperate, because a tit-for-tat strategy is more rational than defection. This result, however, depends on the repetition being open ended. If the parties are aware of when their interactions will end there is a problem of backward induction: neither party has an incentive to cooperate on the n th play, and that means the other party has no incentive to cooperate on the $n - 1$ st play, and so on. For more on the problem of backward induction, see Phillip Pettit and Robert Sugden, "The Backward Induction Paradox," *Journal of Philosophy* 86 (1989): 169–82.

3. Edward F. McClennen, *Rationality and Dynamic Choice: Foundational Explorations* (Cambridge: Cambridge University Press, 1990), and "Pragmatic Rationality and Rules," *Philosophy and Public Affairs* 26 (1997): 210–58; David Gauthier, "Resolute Choice and Rational Deliberation: A Critique and a Defense," *Noûs* 31 (1997): 1–25, "Assure and Threaten," *Ethics* 104 (1991): 690–721.

Pragmatists differ in their approaches to practical reasoning.⁴ I have compared these differing accounts elsewhere, and I will not repeat that discussion here.⁵ Instead I wish to focus on one approach to the pragmatic account, namely, that of David Gauthier. For Gauthier, the form of practical reasoning that accompanies his commitment to the pragmatic approach is constrained maximization: the pragmatic approach to rationality dictates the plans it is optimal to adopt, and constrained maximization is the form of reasoning that allows agents to implement them. Because human beings have the ability to constrain their maximizing, it is rational for them to adopt optimal plans, even when those plans require them to perform actions that are not, considered in and of themselves, maximizing.

Gauthier has famously endorsed the rationality of constrained maximization, but that thesis has taken different forms at different points in his career. In particular, the shift from the dispositional view in *Morals by Agreement* to the plan-based view in “Assure and Threaten” and other post-*Morals by Agreement* papers was an important evolution in his thinking about practical rationality.⁶ In his article for the present symposium, we have a further development: the move from constrained maximization as the basis for implementing rational plans generally to direct optimization in the context of interpersonal coordination. The aspect of Gauthier’s theory I wish to discuss is, as far as I can tell, unaffected by the transition from constrained maximization to optimization. It would not be relevant on the dispositional account of constrained maximization proposed in *Morals by Agreement*, as I shall explain. But since I take that version of the theory to be thoroughly abandoned, and for good reasons, I regard the topic as an abiding concern for Gauthier’s mature view of practical reason.

The subject on which I shall focus is the pragmatist’s ability to cope with plans involving an element of risk and, in particular, the Gauthierian pragmatist’s difficulty with this topic. As will become clear, the features of Gauthier’s account that produce the problem with risk are common to all defensible pragmatic accounts. The only accounts that are both pragmatic and do not suffer from difficulty accounting for risky plans are indefensible for other reasons, comparable to the drawbacks inherent in the dispositional view of practical reasoning spelled out in *Morals by Agreement*. Therefore the only defensible form of pragmatism is that which Gauthier defends in his post-*Morals by Agreement* papers. But *that* version of the theory has a fundamental problem with plans containing elements of risk. While the best versions of pragmatic theories of rationality have a promising solution to the difficulties of expected utility theory, their attractiveness

4. Proponents include David Gauthier, Ned McClennen, Joe Mintoff, Gregory Kavka, Michael Bratman, Scott Shapiro.

5. Claire Finkelstein, “Acting on an Intention,” in *Reason and Intentions*, ed. Bruno Verbeek (Burlington, VT: Ashgate, 2008).

6. David Gauthier, *Morals by Agreement* (Oxford: Oxford University Press, 1986).

is limited by their difficulties accounting for the rationality of plans where the benefits under the plan are less than certain to accrue.

In the literature on the pragmatic account, there is only one sustained discussion of this problem of which I am aware, and that is Gauthier's account of the rationality of plans involving deterrent threats in "Assure and Threaten."⁷ Gauthier noticed a difference in the rational characteristics of threats and assurances, one that seemed to make the gains from cooperation inapplicable to deterrent threats. Several solutions to this problem appear in the literature, but none seems to solve the difficulty. This in turn casts doubt on the pragmatic theory, given that it appears to be unable to vindicate the rationality of plans involving deterrent threats. In this article, I shall suggest that the difficulty Gauthier noticed with plans involving deterrent threats generalizes to any plan involving less than certain benefits. Given the failure of the various solutions Gauthier has explored to that problem, the element of chance in a plan poses a hurdle to the pragmatic account.

In the second half of the article I explore the suggestion that the pragmatist may be able to make his account immune to the asymmetry between sure plans and gambles if he adopts a certain thesis about benefit: the thesis that an *ex ante* chance of benefit is itself a benefit. I call this the Chance Benefit Thesis. I then make a stronger claim: the pragmatist's only reasonable hope for solving the problem of risky plans lies in the plausibility of the Chance Benefit Thesis. After considering several arguments for and against the Chance Benefit Thesis, I reach a weak conclusion in its favor, despite remaining concerns about its plausibility. The strongest conclusion of the article, however, is that if the Chance Benefit Thesis turns out to be indefensible, the objection to the pragmatic account I present here will constitute a sufficient basis for its rejection.

II. PRAGMATIC RATIONALITY AND THE DELIBERATIVE REQUIREMENT

The expected utility theorist denies that it is rational for Bertram to help Alfred plow his field, given that Alfred cannot provide any assurance to Bertram that he will reciprocate. The pragmatist disagrees. Assume, he argues, that Alfred and Bertram are both rational, and that each knows the other is rational. Then it cannot be rational for Alfred to agree to reciprocate but plan not to. For if this were rational, Bertram would know this and would not cooperate with Alfred. The common knowledge assumption effectively rules out asymmetrical solutions; it makes it impossible for Alfred to take advantage of a course of action without Bertram also knowing that Alfred would adopt it. The pragmatist argues, therefore, that

7. See Gauthier, "Assure and Threaten," 709–17.

if common knowledge obtains, there are only two feasible outcomes in reciprocation cases such as these: both cooperate or neither does. And given that Alfred can expect to do better with cooperation than without it, it is rational for Alfred to cooperate and, hence, rational for Bertram to count on Alfred's cooperation.

Some responses to this sort of case seek to vindicate the rationality of performing suboptimal actions in satisfaction of optimal plans by suggesting that it is sometimes rational to act irrationally. Gauthier himself defended this view at one time.⁸ Such accounts are manifestly misguided, however, and they have few adherents now. What we want instead is a defense of the claim that suboptimal actions that are part of, though not instrumentally related to, optimal plans may be rational to perform, rather than rationally-motivated irrationality. How might such a view be defended?

One way to defend the rationality of suboptimal acts is to think of rational plan execution as a two-step process, one that is dependent not only on the agent's formation of an optimal plan, but on the agent's self-motivated execution of that plan on the basis of his reasons for acting. Unlike the earlier accounts that modeled suboptimal actions as irrational, this approach would treat both plan formation and plan execution as requiring rational justification. Rational agents must have an all-things-considered reason not only for adopting the plan in the first place but also for carrying through with its dictates. Whether or not the agent reconsiders a plan is irrelevant, on this view. If an action is rationally justified, reconsideration should not lead to a change of plans, assuming the situation is as the agent expected it to be. For this reason, it should not be necessary in the correct theory of plan adoption and execution to posit a mechanism (psychological or other) to block reconsideration. A rational agent who reconsiders has reason to implement her plan: she has an argument to defeat temptations when reconsideration threatens to lead her astray.

On this view, plan execution is a thoroughly deliberative affair, as subject to the agent's decision making as the initial decision to form the intention or adopt the plan in the first place. I refer elsewhere to the requirement that a theory of rationality explain the performance of rational actions in a way that appeals to the reason or deliberation of the agent the "deliberative" requirement.⁹ I shall adopt that terminology here. The deliberative requirement rules out external precommitment mechanisms. As in the case of Ulysses who tied himself to the mast, it will also rule out what we might call "semiautomatic" devices a rational agent might use in order to mirror the effects of precommitment, but without the external

8. David Gauthier, "Rationality and the Rational Aim," in *Reading Parfit*, ed. Jonathan Dancy (Oxford: Blackwell, 1997).

9. I spell out some of these consequences in "Is Risk a Harm?" *University of Pennsylvania Law Review* 151 (2003): 963–1001, 963.

restraint. Semiautomatic mechanisms include habits, cooperative or other sorts of dispositions, internal resolution that operates solely by blocking reconsideration, rationally induced endogenous preference changes, and so forth. On such accounts, the move that blocks reconsideration is non-deliberative, usually in a semiautomatic way, and therefore plan execution must be a separate matter from plan adoption, which clearly cannot be nondeliberative. But once one has a fully deliberative account of nonreconsideration, the need for a separate theory itself drops out. What the agent needs is a rational basis for proceeding with the prior plan he or she has already adopted.

Once an agent has a rational basis for acting of this sort, it does not particularly matter whether she reconsiders: if she does, she will decide in favor of proceeding with the previously formed intention for precisely the same reason she adopted the plan in the first place. What this means for reciprocation cases is that we need an account that can explain how farmer Alfred can have a reason for actually plowing Bertram's field. It is not necessary to explain the rationality of not reconsidering his intention to plow Bertram's field once formed.

The deliberative requirement does not place any constraints on the kind of reason that will qualify in this regard; it merely says that there must be some such reason. But insofar as the pragmatist is articulating a theory of rationality, there are restrictions beyond the deliberative requirement, restrictions about what kinds of reason could count as rationalizing Alfred's plowing of Bertram's field. A theory of rationality, for example, cannot appeal to moral reasons for this purpose, though moral reasons would satisfy the deliberative requirement. A further restriction on the kinds of explanations a pragmatist can offer, then, is that the reasons to which agents appeal must be reasons of self-interest. It may be that in following reasons of self-interest, the pragmatically motivated agent will behave morally as well, insofar as his reason would lead him to keep promises, to cooperate with others, and so forth. But the conformity to moral norms would then be a side benefit of adherence to the pragmatic theory of rationality. It would not itself be a reason for adopting such a theory.

Proceeding against the background of both the deliberative requirement and the maximizing conception of rationality, there would appear to be only one version of the pragmatic theory that fits our desiderata: the account of rationality for which Gauthier argues in his essays post-*Morals by Agreement*.¹⁰ In those papers, Gauthier argues that an agent should conform to the dictates of a plan if and only if by her lights she is better off under the plan than she would have been had she never adopted the plan at all. An agent who follows this principle executes her plans rationally,

10. These papers constitute a rejection of the *Morals by Agreement* account. David Gauthier, "Intention and Deliberation," in *Modeling Rationality, Morality and Evolution*, ed. P. Danielson

meaning that there is a reason that guides her deliberative faculties during both plan execution and plan formation. The theory therefore satisfies the deliberative requirement, and the content of the reason satisfies the rationality constraint. Indeed, this test is the only way a pragmatist can vindicate the rationality of performing suboptimal actions in the kind of case we are considering, consistent with the deliberate requirement. For given the basic assumption that human beings are rational maximizers, the deliberative requirement can only be satisfied if, in executing a plan, the agent can see herself as better off under the plan than she would have been in its absence. Gauthier calls a plan that satisfies this condition “fully confirmed.”

Full confirmation is thus the practical rationality equivalent of an agent’s being disposed in the *Morals by Agreement* model to cooperate with others, namely, having the disposition of a constrained maximizer. In the *Morals by Agreement* model, constraining one’s maximizing in accordance with optimal plans is rational, even if this requires disposing oneself to perform actions that are not, in and of themselves, maximizing. Reaping the gains of cooperation in reciprocation games and prisoner’s dilemmas cannot be achieved without forming certain kinds of dispositions under this model. But actions performed on the basis of dispositions are not deliberatively based and thus fail to satisfy the deliberative requirement. Gauthier’s post-*Morals by Agreement* notion of full confirmation, by contrast, does satisfy the deliberative principle, and this gives us a seemingly attractive way of cashing out the optimality condition the pragmatist endorses.¹¹ I shall accordingly say that the post-*Morals by Agreement* approach to constrained maximizing, based as it is on reflective execution of optimal plans, satisfies the Pragmatic Deliberative Principle.

According to the Pragmatic Deliberative Principle, when the time comes for Alfred to decide whether to make good on his promise, he should ask himself how he would have fared had he never entered into the agreement with Bertram in the first place. He should compare this to the position he is in under the terms of the agreement, counting the costs of compliance. Without the agreement, he would be left to plow his field by himself, but he would not have to plow Bertram’s field. With the

(Oxford: Oxford University Press, 1998), 41–54, “Resolute Choice and Rational Deliberation: A Critique and a Defence,” *Noûs* 31 (1997): 1–25, 20, “Commitment and Choice: An Essay on the Rationality of Plans,” in *Ethics, Rationality and Economic Behavior*, ed. Francesco Farina, Frank Hahn, and Stefano Vannucci (Oxford: Clarendon, 1996), 217–43, “Individual Reason,” in *Reason, Ethics and Society*, ed. J. B. Schneewind (Chicago: Open Court Press, 1996), 39–57, and “Assure and Threaten.”

11. Gauthier says a plan is “confirmed” at a given time, “if at that time the agent may reasonably expect to do better continuing it than she would have expected to do had she not adopted it.” See Gauthier, “Intention and Deliberation,” 49.

agreement, he has the cost of plowing Bertram's field, along with the benefit of Bertram's assistance. When he compares his life under the plan with his life in the absence of the plan, he sees he is better off with the plan than without it. Under these conditions, it is rational to follow through with the plan.

There are many objections to the pragmatic account in the philosophical literature, and many answers to those objections. It is not my purpose, however, to offer a general defense of the pragmatic account. My aim instead is to highlight a very particular problem with the pragmatic account, one that threatens to make any victory over expected utility theory a pyrrhic one, namely, that the pragmatist has difficulty defending the merits of plan execution where the benefits from the plan are less than fully certain to accrue. Some pragmatic accounts appear to address this problem effectively, but to date all such accounts have either implicitly or explicitly assumed a nondeliberative method of plan execution. My point can thus be put as follows: once the deliberative requirement is fully accommodated, it becomes difficult for the pragmatist to deal effectively with risky plans without turning back to standard expected utility theory. This poses a serious challenge to the pragmatic account.

III. RISKY ASSURANCES

Sometimes the optimal plan is one that requires an agent to gamble. No matter how risk averse the agent, there will be some plans involving gambles whose expected benefits are sufficiently high and whose risks are sufficiently low, that a rational agent would regard the plan as optimal. The problem for the pragmatist is that if the gamble fails, the actor will have to perform suboptimal acts required by a plan knowing that the benefits she hoped for in adopting the plan will never accrue. In such cases, the pragmatist will not be able to vindicate the rationality of particular acts by turning to the benefits of the plan, since the plan will not in fact have left the agent better off than she would have been had she never adopted the plan in the first place. For a rational agent following the Pragmatic Deliberative Principle, it cannot be rational to stick to the plan in the case in which the agent loses the gamble, since this would make the agent's life go worse, rather than better overall.¹² If following a plan would make the agent's life go worse, she cannot rationally justify performing any suboptimal act required by it, since there can be no reason for performing such acts—neither the plan nor the payoff from the act itself supplies one.

12. Notice that the economist has no particular difficulty with risky plans. For his position is that it is not rational to implement a plan that calls for a suboptimal act if the latter is not a means to the promised gains from the plan, and that if the act is a means, the act is rational. In no event does it matter whether the act must be performed in the face of

Suppose the situation the farmers face is like this. Alfred's field is twice as large as Bertram's. Alfred, therefore, proposes that Bertram give him, Alfred, a .5 chance of receiving Bertram's assistance, in exchange for a definite commitment on Alfred's part to plow Bertram's field next week. Let us assume Alfred does not have any particular preference with regard to risk; he is risk neutral. Where does the rationality of their agreement stand in this case? On the one hand, both parties should regard this deal as advantageous if the original proposal was, since each has the same expected utility he had under the original set of circumstances. But suppose Alfred flips a coin to determine whether Bertram will help plow his field, and he loses. Under the terms of the agreement, Bertram will not help Alfred plow his field, but Alfred must still plow Bertram's next week. Is it rational for Alfred to follow through with the plan and proceed to plow Bertram's field as promised?

According to the Pragmatic Deliberative Principle, it is not. For when Alfred now asks himself whether his life will go better under the plan than it would have gone had he never adopted the plan in the first place, he will have to answer the counterfactual test in the negative. Alfred now desperately wishes he had not entered into the agreement, for it has turned out to be all cost and no gain to him. And given common knowledge of rationality, Bertram knows it is not rational for Alfred to reciprocate should he lose the gamble. The agreement is not rational for Bertram either. The Pragmatic Deliberative Principle thus seems to fail as applied to reciprocity cases, at least if the assurance involves less than certain payoffs.

In "Assure and Threaten," Gauthier partially notices the pragmatist's difficulty with risk, but only in the context of threats.¹³ What drives the article is the concern that under the pragmatic theory, it cannot be rational for a person to threaten another to deter him from doing something if making good on the threat would be costly for the person issuing the threat. Should the threat fail to deter, he cannot (rationally) carry out his threat, since the threat will turn out to have been all cost and no gain. But given the possibility of risky assurances we have identified, the problem applies equally to assurances. One way to put the point is in terms of the deliberative requirement: plans involving deterrent threats and risky assurances cannot be rational to adopt if they must be deliberatively executed, since a rational agent cannot know whether her life will in fact go better under the plan.

certain or uncertain gains. The economist will thus endorse entering into certain gambles but will see no reason to make good on those gambles in the absence of some precommitment mechanism. Whether one wins or loses a gamble is irrelevant to the rationality of carrying on as the plan requires.

13. See Gauthier, "Assure and Threaten," 709–13.

This seems to suggest that the only way to vindicate the rationality of plans involving risk is in expected benefit terms, namely, in terms of the increased chances of achieving the sought for benefit. But expected benefit calculations will not rationalize performing a costly action in fulfillment of a plan that must be deliberately performed, if the action does not have autonomous benefits. So expected benefit based justifications require automatic or semiautomatic plan execution. Adoption of automatic or semiautomatic methods of plan execution is not consistent with the deliberative requirement—a fundamental condition of rational agency.

Thus far I have stuck with what I take to be Gauthier's original approach to resolute choice, namely, the one that most immediately replaced the dispositional account of *Morals by Agreement*. But we have now to bring on board his attempted solution to the problem of the rationality of plans involving deterrent threats in "Assure and Threaten." If it had succeeded, Gauthier's only mistake would have been his failure to extend that solution to the problem of risky assurances. But as I have already suggested, the solution appears to fail. The difficulties here will be instructive.

IV. THE POLICY APPROACH

On Gauthier's approach, it is irrational to issue threats that would be costly for the issuer to execute. It might nevertheless be rational to have a policy of issuing and making good on such threats. While no single threat is certain to pay off, a policy of threat issuance and execution seems more promising. Having such a policy, after all, allows an agent to use threat issuance as an effective means of convincing others to adhere to her demands. The original pragmatic deliberative approach would thus require only a slight modification: rational agents should select actions in accordance with optimal policies instead of with optimal plans. In order to determine whether to adopt a given policy, the rational agent should use the familiar counterfactual test: he should ask himself whether he is better off under the policy overall than he would be in its absence, taking into account the costs of making good on a particular deterrent threat. If the answer is "yes," then it is rational for the agent to follow through on a plan involving that threat, since making good on the threat constitutes the execution of a policy of threat issuance and threat execution it was rational to adopt.

Unfortunately, this move to a higher level of generality does not solve the difficulty with deliberative execution of suboptimal actions. Just as the appeal to plans unravels when rational agents are confronted with a finite number of iterations of cooperative behavior, so moving to the higher level of abstraction involved in reasoning from policy revives the problem we saw in the context of plans: if a given mode of reasoning say, reasoning from policies, would require me to do something that is not itself utility

maximizing, it is rational to reject it. But just as David Lyons showed for the relation of act to rule utilitarianism,¹⁴ it is always possible to alter my mode of reasoning just slightly to incorporate exceptions to a general policy, thus forming a new, improved policy that is better contextualized to maximize utility. For this reason it cannot be rational to adopt a mode of reasoning that recommends suboptimal acts. Alternatively we should be able to trade in one policy for a “better” policy from the standpoint of expected utility theory—one that alters just the suboptimal act in question.

There is a second problem that is specific to the policy approach Gauthier proposes in “Assure and Threaten,” one I think Gauthier himself has accepted as fatal to any attempt to rationalize deterrent threats on the basis of reasoning from optimal policies.¹⁵ When a rational agent decides to adopt a policy of threat issuance and threat execution, she cannot be certain that implementing the policy will in fact make her life go better than not adopting the policy at all. Though adherence to a policy of following through on deterrent threats is likely to result in the issuance of effective deterrent threats, it might turn out that no threat ever actually succeeds in deterring, given that the recipient of these threats might always resist. Adopting the policy would then leave the agent worse off than had she not adopted it at all.¹⁶ If we consider the rationality of a threat policy *ex ante*, we cannot assume it is rational to adhere to a policy of making good on deterrent threats.

Third, there is a problem of “backward induction” that will plague any effort to account for threat execution in terms of policies. Suppose we know how many instances of threat issuance and execution we will encounter under a given policy. Then on the last instance of the series, a given individual will have no reason to make good on his deterrent threat, because there will be nothing further to be gained from his doing so, and following through on the threat will no longer undermine the policy of threat issuance and execution that seems advantageous when in the midst of the policy. Following through on the last threat will be all cost and no gain. So in a policy with n threats, it makes sense to stop executing threats issued on the $n - 1$ threat. If this is true, however, then it is not rational to issue the n th threat, since against the background of common knowledge of rationality, the recipient of the n th threat will know it is not rational to execute that threat, and so will not be deterred. But now that we have dis-

14. David Lyons, *Forms and Limits of Utilitarianism* (New York: Oxford University Press, 1965).

15. Unpublished exchange with David Gauthier. The argument is one originally noted by Joe Mintoff, in a correspondence he had many years ago with Gauthier after the publication of “Assure and Threaten.”

16. True, if an agent suddenly happens to find herself in the middle of a policy which has already proven beneficial, she can appeal to the benefits of a policy in order to justify making good on a particular threat.

pensed with the rationality of making good on the n th threat, we will encounter the same problem with respect to the $n - 1$ threat, and then with the $n - 2$ threat, and so on down the line. Thus we can eliminate all rational threat execution under the policy by backward induction.

A defense against the backward induction argument would be that we do not normally know how many instances of threat issuance and execution would be necessary under a rational threat policy. It is more realistic to suppose that any such policy will be open ended. But it is not necessary to know which of the threats I issue will be the last one for the backward induction objection to apply. It is enough to know there will be a last threat for the problem to arise: if I know that the number of threats is finite, I know that the last threat, whenever it is, will not be rational to issue, as well as the threat before that, back to the first threat. Thus the mere knowledge that the number of threats is finite will start the backward induction, and eliminate the rationality of appealing to a policy of threat issuance and execution. Since the finitude of any approach to practical reasoning can be inferred from our mortality, it appears that the appeal to policy will fail.¹⁷

For the foregoing reasons, then, the move from rational plan adoption and execution to rational policy adoption does not appear to solve the difficulties raised by the rationality of following through on failed deterrent threats. And given the parallel between deterrent threats and risky assurances, the turn to policy will not help to rationalize the issuance and execution of risky assurances either. So the problem with the rationality of threats Gauthier identified in "Assure and Threaten" is indeed serious, and we are left without a solution.

V. THE CHANCE BENEFIT SOLUTION

Because of the backward induction problem discussed in the preceding section, Gauthier eventually concluded that the notion of a policy cannot vindicate the rationality of plans involving deterrent threats. That conclusion seems fundamentally sound against the background of Gauthier's approach to rationality: plans involving deterrent threats may require the person issuing the threat to perform a suboptimal act—make good on the threat—if the threat fails to deter. Not only is following through on the threat suboptimal, as it is with assurances, but the plan as a whole will have failed to yield any benefit. Thus unlike for assurances, where de-

17. It might be argued that the relevant knowledge is not the finitude of the chain of threats, but knowing, for any given threat, that it is the last threat. But this seems incorrect. For given the backward induction, I know not only that it is irrational for me to issue the last threat, whenever it is, and that this makes it irrational for me to issue the $n - 1$ threat, and the $n - 2$ threat, and so on, then I know that wherever the threat I am about to issue falls in the series, it is irrational to issue.

terrent threats are concerned, it is not possible to vindicate the rationality of otherwise suboptimal actions by reference to the larger plan of which they are a part.

It is not surprising, then, that embedding the suboptimal plan in a still more general entity—a policy of threat issuance and threat execution—does not improve matters. Just as the plan may turn out to be suboptimal if the threat fails to deter, so the policy may turn out to be suboptimal if the policy fails to deter.

As we have seen, if dispositions supply the relevant “mode of reasoning,” instead of policies, matters are otherwise: since dispositions are largely self-executing, acts done on the basis of dispositions need not have positive payoffs for rational agents to perform them. Actions performed on the basis of dispositions, however, do not satisfy the deliberative requirement.¹⁸ They therefore do not count as “rational actions” in the relevant sense and so cannot solve the problem of the rationality of deterrent threats or risky assurances either.

Let us then return to the counterfactual test as Gauthier originally formulated it and ask whether there are other possible solutions to the problem posed by threats and risky assurances. Recall the original problem we faced with risky assurances. Suppose the payoffs for Alfred are as follows: receive help and render no help (16), exchange help (12), receive no help and render no help (5), render help and receive no help (1). Alfred’s baseline (his payoff in the absence of any agreement) is 5, because that is the payoff from defecting from the cooperative scheme (render no help), which would imply that the agent also receives no help. In the game tree depicted in figure 1, only Alfred’s utilities are represented. Call “receiving no help” renegeing. We treat the first node of the game tree as a chance node, because whether Bertram cooperates or defects from their agreement is not under Alfred’s control. Therefore we treat Alfred as “winning” if Bertram reciprocates, and as losing if he does not. We can then model the payoffs for Alfred as shown in figure 1.

Alfred’s preferred outcome is of course to receive help and to renege, since it would give him a payoff of 16. But this outcome is not available to him, since against the background of common knowledge of rationality, we can assume that he will not receive help if it is rational for him to renege after he has done so. So the next best available outcome would be for Alfred to win the gamble and to make good on his promise by cooperating with Bertram (for a payoff of 12). Since the position he would be in were he to forgo the agreement altogether is 5, he can tell himself he is

18. At least this is the case on Gauthier’s account of dispositions. While it might be possible to hold a view of dispositions that made them compatible with deliberation, such an account would encounter the same difficulty with the rational justification for the acts recommended by the disposition that we saw with policies and plans.

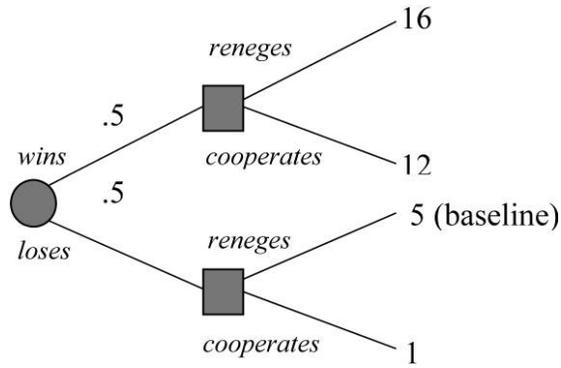


FIG. 1

better off than he would have been had he never agreed to reciprocate. The problem is that if he loses the gamble, making good on his promise would leave him worse off than he would have been without the agreement, since he would end up with a payoff of 1. In this case, the agreement to cooperate with Bertram does not satisfy the pragmatic deliberative principle, and it would not be rational for him to reciprocate. Since Alfred knows this in advance, he cannot commit to following through should he lose the gamble. The result is that he cannot sincerely promise to reciprocate, and Bertram cannot trust Alfred to reciprocate. Bertram will refuse to plow Alfred’s field.

On Gauthier’s account, reciprocation is not rational unless Alfred can see himself as benefited by the agreement with Bertram. But how can he see himself as benefited if he is on the losing end of a risky assurance or deterrent threat? It seems he cannot. The only other option we have seen so far is for the plan to be at least partially self-executing, as it might be if it were to be executed on the basis of dispositions. But the account would then fail the deliberative requirement, and arguments in favor of retaining that requirement are strong.

There is, however, a possibility we have not yet considered, one that makes use of the notion of chance benefit we earlier defined. If there is such a thing as chance benefit, a person who loses a gamble may still have been benefited: she has been benefited by the chance of benefit the gamble afforded her. Thus if the .5 chance farmer Alfred had to secure Bertram’s assistance is a benefit in and of itself, this would enable him in some cases of risky assurances to still see himself as better off under the agreement with Bertram, come what may. And this is true even if Alfred loses the gamble. Chance benefit is a thin gain as compared with actually winning the gamble outright—a notion we will refer to as outcome benefit. Still, if the chance benefit is large enough, Bertram’s welfare might be improved, relative to his starting position, on the losing branch of the gamble.

But isn't chance benefit just a chance of receiving some quantum of outcome benefit? And if so, how could it do any independent work in defining the payoffs of a gamble? Instead of rejecting this idea out of hand, let us consider what would have to be the case for chance benefit to function in this way. For chance benefit to have value in the way we are exploring, it must persist when outcome benefit fails to materialize. We might more naturally, however, think of chance benefit as absorbed into the outcome benefit when a person wins the gamble. To have a basis for rejecting that absorption of chance benefit into outcome benefit, we would have to think of the chance of winning as itself a benefit. Is this plausible?

The idea seems perhaps less far-fetched when one considers the same kind of idea with regard to harm and risk. Let us call the notion that corresponds to chance benefit on the harm side *risk harm*. The person who wins a gamble has been exposed to a risk of loss, and that exposure is itself a kind of loss. Her gains from winning the gamble are therefore reduced as a function of the risk she ran of losing rather than winning. Just as chance benefit is absorbed into outcome benefit when the chance of benefit eventuates, so the risk of losing is absorbed into the loss when the risk eventuates in outcome loss. On this view, the person who wins a gamble is somewhat worse off than his outcome winnings would suggest, due to the risk of losing to which he was exposed, but the person who loses is not doubly worse off because in addition to losing, he ran a risk of losing.

I have not attempted anything like a full-fledged defense of the notions of chance benefit and risk harm. But it is nevertheless interesting to note the implications for the Gauthierian pragmatist if these concepts were viable. Where the chance benefit is large, it would allow a person to see herself as benefited in some cases, even where she has not secured any outcome benefit from a given assurance or threat. Those risky plans would be rational to adopt when the *ex ante* chance of benefit is large enough to outweigh the *ex post* costs of making good on the plan. This would allow the pragmatist to say that some risky assurances and some deterrent threats are rational to issue, since in some cases, even if the threat fails to deter, it would be rational to make good on the threat in light of the chance benefit that issuing the threat provides. The plan would thus satisfy the Pragmatic Deliberative Principle. It would also require some adjustment on the winning side, as emerges in the case of assurances. Some assurances that appear to be rational based on the expected benefit of the winnings will turn out not to be worthwhile, once we factor in the risk harm, since the agent's gains must be reduced by the amount of risk harm to which she is exposed along the way to winning.

Now admittedly it is not easy to see how we would measure such a thing as chance benefit or risk harm. For one might suppose that the amount an agent values a chance at receiving a certain outcome benefit will depend at least in part on highly individualistic reactions agents have

towards risk. A risk-loving agent will place a disproportionate value on chance benefit, since in addition to valuing the chance of receiving a certain outcome benefit, he will also place a positive value on the element of risk itself. A risk-averse agent, by contrast, may undervalue a risky assurance as compared with the expected benefit of the gamble. She would presumably discount the expected benefit by the amount of risk the gamble involves, since for her, exposure to risk is in itself a negative feature of the situation. But on a nonepistemic approach to the concepts of chance of benefit and risk of harm, there is an objective measure of benefit an agent receives from a chance of outcome benefit, whether she values that chance correctly or not. And it is this feature that may allow us to think of chances and risks as affecting the ultimate value of plans with probabilistic elements. For the sake of simplicity, let us treat all agents as risk-neutral agents. For risk-neutral agents, the chance benefit of a gamble is roughly the value of the benefit she hopes to get out of the gamble, in outcome terms, discounted by the risk of not getting it.

In order to see how the pragmatic account might be modified in light of the notion of chance benefit, let us alter the simple model of an assurance we have just considered. Suppose once again that Bertram presents Alfred with a .5 chance of assistance. The chance benefit for the agent, assuming risk neutrality, is the value to him of the increase in outcome benefit he hopes for under the plan, times the probability of receiving these benefits. In order to test the rationality of the plan, we must consider the worst payoff the agent could have under it, adding together the outcome benefit and the chance benefit she could acquire. We should then consider whether, taking both kinds of benefit into account, the agent would be better off than if she had never adopted the plan at all. Call the "no plan" payoff the agent's "baseline."

In this case, the worst scenario for Alfred under the plan would occur if he were to lose the gamble and proceed to cooperate nonetheless. His outcome payoff would be 1. His chance benefit, however, would be the chance he had of receiving an increased payoff of 12 over and above his baseline payoff. The baseline in this case is equal to the payoff from losing the gamble and then renegeing, since Alfred's payoff under that scenario is the same as his payoff if neither assists the other, and this is the outcome if no plan is made. Since his increase would be 7, and he has a .5 chance of receiving that increase, he has a chance benefit of 3.5. If we aggregate chance benefit and outcome benefit on the losing side, his total payoff at the worst outcome is 4.5. We then compare this to how Alfred would have fared had he never entered into the agreement with Bertram in the first place, and we see that Alfred still fares worse under the plan in the worst case scenario than he would have fared otherwise, since without the plan he would have ended up at 5, and with the plan he will end up at 4.5 if he loses the gamble. If we calculate the risk harm on

the winning side, we must subtract 1 from his baseline welfare of 5 or 4, offset by a .5 chance of non-occurrence, to represent a risk harm of 2. The winning scenario under the plan would therefore give a combined payoff of 10 (see fig. 2).¹⁹

But suppose now that the payoffs from winning the gamble are higher. Let us suppose that if he wins the gamble and proceeds to cooperate, Alfred's payoff would instead be 20 (and suppose his payoff from renegeing, were that option available, would be 24), and that if he does not enter into the agreement, his utility is still 5 and his outcome payoff is 1 if he loses the gamble. Then the chance benefit of the plan would be $.5 \times (20 - 5)$, or 7.5. Adding the chance benefit to an outcome benefit of 1 should he lose the gamble, Alfred would end up at 8.5 if he loses the gamble. Since that is better than the 5 he would obtain if he does not enter into the plan in the first place, he would be better off gambling, even if he loses and has to reciprocate without tangible outcome benefit from the deal, than if he had never agreed to cooperate with Bertram in the first place.

To be clear about how chance benefit and risk harm work, let us also consider the value of the winning side of the gamble. There Alfred would have an outcome benefit payoff of 20, but the payoff must be reduced by the amount of risk harm to which he has been exposed. That risk is a .5 chance of having to help without receiving help, which would set Alfred back by 4 as compared with where he would be had he never entered into the agreement in the first place. So the risk harm to Alfred is $.5 \times 4$, or 2. The risk harm would thus clearly represent a cost to Alfred, since it is negative relative to Alfred's baseline. We then reduce the cooperative payoff of 20 by the risk harm of 2, to get a total payoff of 18. Because under this plan, the agent comes out ahead whether he wins or loses, the plan is rational to adopt. This is one scenario, then, where a plan that is not "fully confirmed," in terms of outcome benefit, passes that test on the chance benefit model I have suggested.

Where threats are concerned, the analysis is somewhat trickier, but the basic account is the same. Suppose instead of cooperating, Alfred and Bertram are in a different situation. Alfred wants to deter Bertram from applying for a job he hopes to obtain, and so Alfred considers threatening to tell Bertram's wife that Bertram is having an affair. Suppose further that if Alfred does threaten Bertram, his threat has a .6 chance of deterring Bertram from applying for the job, and a .4 chance he will be undeterred and will apply for the job anyway. Alfred's payoff, should he deter

19. Note that we do not have to calculate the chances that we will choose to cooperate or that we will choose to renege. In this case, the agent is considering the rationality of the plan that involves cooperation. We therefore assume, *ex ante*, that the agent will choose cooperate, rather than renege, when he comes to the choice node.

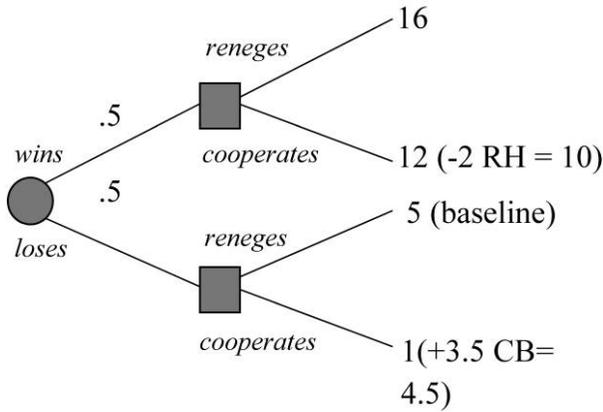


FIG. 2

Bertram, would be 16. If Alfred fails to deter Bertram, however, and should Alfred then make good on his threat, he would be left with a payoff of 0. Suppose also that if he were to defect from his plan of making good on his threat, Alfred would have a payoff of 5. This is also his payoff if he had never threatened Bertram in the first place and Bertram proceeded to apply for the job anyway. Finally, there is a small chance (.2) that Bertram will not end up applying for the job, even if Alfred doesn't threaten him, in which case Alfred's payoff will once again be 16. The decision tree for this plan looks like the representation in figure 3.

Recall that the chance benefit is the amount of potential increased outcome benefit an agent has over and above his baseline welfare (his well-being without the plan), multiplied by the probability of his receiving that increase. The increased payoff for which he hopes under the plan is 16—the payoff if deterrence succeeds. Here the no-interaction baseline is spread over two possible outcomes, which we must aggregate. It is $.8(5) + .2(16) = 4 + 3.2 = 7.2$. So the chance benefit from the plan is $16 - 7.2 = 8.8$. We must now add that to the lowest payoff under the plan (zero) to see if the plan is worthwhile. Since the smallest aggregate benefit Alfred would receive under the plan is 5.28, and it is substantially lower than his baseline of 7.2, the plan in this case is not rational to adopt.

If we wish to calculate the winning branch of the tree, we must subtract the risk harm from the outcome payoff of 16. The risk harm is the chance Alfred has of suffering a loss below his baseline. That baseline, as we said, was 7.2, and he has a .4 chance of losing the entire amount. We therefore calculate the risk harm as $7.2 (.4) = 2.88$, and we subtract this from 16 to get 13.12. This shows that the plan would be worthwhile on the basis of the winning branch, but not on the basis of the losing one. Since the deliberative requirement demands that the plan provide a net benefit under

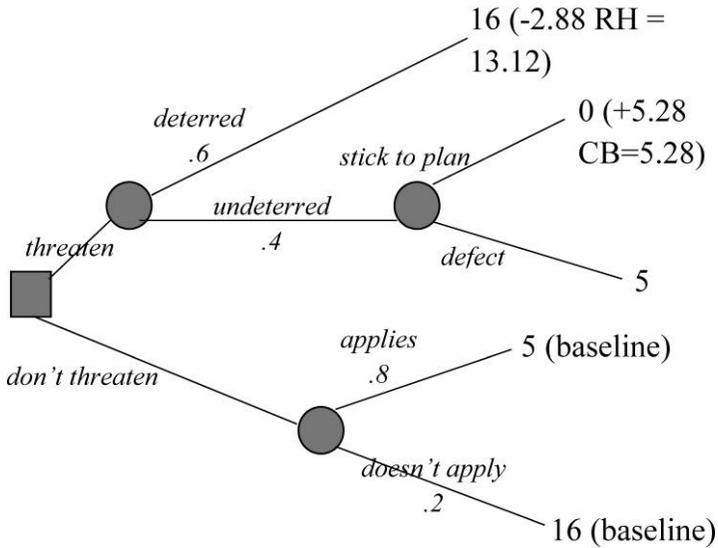


FIG. 3

every reasonably likely outcome, the plan would not be rational to adopt. Plans involving deterrent threats might be “fully confirmed,” as Gauthier would say, but the instances will be relatively rare. We might thus provide symmetrical accounts of chance benefits and risk harms, in a way that would permit an explanation for why some risky assurances and some deterrent threats are rational to issue. If successful, we would have obviated the need to explain the rationality of suboptimal actions embedded in optimal plans in terms of more generalized modes of reasoning, such as optimal policies.

VI. RESPONSE TO OBJECTIONS

While the chance benefit account may derive some support from intuition, the basic thought on which it depends seems difficult to accept: how can farmer Alfred regard himself as better off plowing Bertram’s field without receiving Bertram’s assistance than he would have been had he never agreed to Bertram’s gamble? Unless an agent is risk loving, there is no reason to suppose she would regard exposure to a gamble as advantageous per se. Nevertheless, I think the notion of chance benefit may not be as unintuitive as it at first glance appears, and that Alfred may have reason to regard himself as better off having had the opportunity for gain, despite the fact that he has lost the wager. Given the objective nature of the concept of probability we are using, Alfred is better off under some risky gambles, whatever his view of it happens to be.

First, despite the fact that we have rejected the idea that the probabilities involved are merely epistemic, it is nevertheless instructive that people value risk in different ways. Some people are risk averse, meaning that they discount the same expected value the greater the element of risk involved. Other people are risk loving, meaning that it is a benefit to them that they have a larger potential payoff which is less than certain to occur, as compared with a smaller certain payoff. These different preferences with respect to risk suggest that people value chances of benefit somewhat distinctly from the way they value beneficial outcomes themselves. In other words, they treat a chance as a separate object of value. Admittedly, this does not prove the point. For even if the chance of a benefit is seen as a separate item of value, this does not mean that its value is independent of outcome benefit. A risk-neutral agent values the chance of a benefit at the expected value of the outcome. But this only means that such a person has no special preferences regarding chances and risks, as compared with outcomes. We cannot know from this whether the person treats a chance of benefit as itself a benefit.

The most plausible way to cash out the notion of chance benefit is in terms of objective notions of frequency. Thus the fact that a person values risk ought not, in principle, to increase the value of a gamble. Most claims that chance benefit collapses into outcome benefit will ultimately depend on epistemic, rather than objective, notions of probability. This for some will be a sufficient basis for rejecting the notions of chance benefit and risk harm.²⁰ But there are many adherents of objective notions of frequency, and while it is beyond the scope of the present paper to argue in favor of objective over epistemic approaches to probability, siding with such an account is not a highly controversial move.

Second, there are a number of examples in which the idea that an *ex ante* chance of benefit is itself a benefit receives support from intuition. Suppose someone gives you a lottery ticket as a present. You would tend to feel that he has benefited you, and you would feel this even if it turned out not to be the winning ticket. True, the ticket has no further value to you once you learn it is not the winning one. But this only means that the benefit that an *ex ante* chance of benefit confers is “parasitic” on outcome benefit. A chance of benefit, however, may still be a separate form of benefit, valued in a somewhat different way. Notice, in support of seeing a lottery ticket as a benefit, that no one would enter actual lotteries if they only valued tickets at their expected value. For the expected value of the ticket is lower than its price. But of course this just shows that gamblers are risk loving, which is hardly a surprise.

20. See Stephen Perry, “Risk, Harm and Responsibility,” in *Philosophical Foundations of Tort Law*, ed. David G. Owen (Oxford: Oxford University Press, 1995), 321–46.

The strongest support for the thesis, however, emerges if we consider harm rather than benefit. For it lends credence to the idea of chance benefit if we can make sense of its converse, risk harm. As it happens, the law is replete with cases in which courts seem to treat a risk of harm as a harm. Mass tort cases have often allowed plaintiffs to recover for an exposure to harm, even if no actual injury results. Someone exposed to an environmental toxin for many years, which has caused others to develop cancer, may be able to recover purely on the basis of exposure to risk. This accordingly allows him to recover at the time of exposure, rather than waiting to see if the risk materializes. The criminalization of reckless driving may be another example, given that it exists alongside a general commitment to restricting the use of the criminal sanction to instances of harm prevention. Of course criminalizing risky activities does help prevent resultant harm. But the same could be accomplished simply by punishing the actual infliction of injury more severely. Still, if we did not perceive imposing a risk as a form of harm infliction, it is not clear we would be justified in prohibiting risk-imposing conduct *per se*.

It may be tempting to think that these cases are actually instances of holding people liable for violating rights, rather than for inflicting harm. If so, they would not support the intuition that we treat imposing a risk as tantamount to inflicting harm but would instead call for a different kind of analysis altogether.²¹ While this thought may have surface appeal, it actually makes little sense. What would be the basis for saying that exposing someone to a risk violates his rights? What would be the basis for the violation? Presumably it violates someone's rights to expose him to a risk of death or serious bodily injury because it threatens to harm him. Should we not then say that exposing someone to Russian roulette without his consent, for example, violates the victim's rights because it threatens to harm him some time in the future, and that, in itself, constitutes a rights violation? It seems more natural to say that threatening to harm someone in the future may constitute threatening to violate his rights at that future time, and that the rights violation and the harm seem to go hand in hand, at least when the harm in question is an impermissible one. Thus it seems more natural to say that the reason exposing someone to a risk of harm violates his rights (when it does) is that it constitutes a setback to his interests now, not just at some point in the future.

Perhaps the most interesting legal cases for our purposes are the medical malpractice cases involving so-called "lost chance of benefit." Traditionally in order for a plaintiff to recover for injuries in a civil case, he must be able to show that it was more likely than not that the defendant caused his injuries. Thus if a patient with a potentially fatal disease was maltreated by

21. I am indebted to an anonymous reviewer for *Ethics* for this point.

a negligent doctor, it might turn out to be difficult for his estate to recover if he died, since the defendant will always say that the patient would have died anyway. Courts have increasingly started to say, however, that a patient with a 40 percent chance of survival, whose chances were reduced to 20 percent by a doctor's negligence, can recover for the lost chance of benefit, valued at 20 percent of the extent of his injuries. Since it is not clear in that case that the doctor caused the plaintiff's injury, such decisions award compensation for the increased risk of death the doctor imposed, or, alternatively, the decreased chance of cure.

There are, however, several difficulties with the concepts of chance benefit and risk harm. One of the biggest problems is the potentially redundant nature of the benefit. If I receive a chance benefit from exposure to the possibility of winning, and that benefit should be added to the value of my position once I have lost, then are we not also compelled to say I have a chance benefit that should be added to my winnings if I were to win the gamble? Would that not mean that we are engaged in a kind of double counting, since the winnings are supposed to be the realization of the chance benefit? If this were the case, the power company that exposes me to risk of cancer would have to compensate me twice in some cases: first they might owe me compensation for exposure to risk, and then if I actually develop cancer, they would also have to compensate me for that. It would follow that the person who developed cancer as the result of a process that was certain to produce cancer would receive less compensation than the person who was first only exposed to a risk and then developed it later. Since we would normally think of the person who developed cancer as the result of a sure process as more seriously wronged, rather than less, this would produce very counterintuitive results indeed.

It was partly in order to avoid results like the foregoing that I claimed that chance benefit and risk harm are asymmetric with respect to gains and losses: the value of chance benefit is absorbed into the winnings if an agent actually wins, and the (dis)value of the risk harm is absorbed into the loss if the agent does indeed lose. It is only in the case in which an agent loses that we should think of the chance benefit as persisting, and the case in which the agent wins that we should think of the risk harm as persisting. But what is to justify this asymmetric treatment of chance benefit and risk harm? The thought does not seem as unintuitive as one might suppose. The idea is that a person who has lost a gamble can regard herself as somewhat better off for having had the shot at winning. Think of a person who has been turned down for a job for which he was seriously considered, as compared with the person who never was formally considered in the first place. But there seems no intuitive support for the claim that a person who actually wins can think of herself as better off than even her actual winnings suggest for having gotten those winnings by chance rather

than by a certain process. That is, the person who actually gets the job surely does not think of himself as better off because he was less than certain to receive it. Similarly, a person who has won might regard herself as in some sense harmed by having been exposed to a risk of harm along the way to winning. Imagine learning upon disembarking from a plane that the airline knowingly flew with only one engine. But it would be exceedingly odd to think of oneself as worse off for having lost by a risky process than by a sure one. The person who is harmed in a car accident presumably does not think of herself as worse off than the person who sustains the same injuries through an intentional battery.

The notions of chance benefit and risk harm, then, treat the expected value to an agent of the other branches in a risky plan as relevant to the value an agent attaches to the branch on which he finds himself. We assume, that is, that an agent is not entirely indifferent to the manner by which she comes about her winnings. But this is just to say that the value agents place on outcomes is in some sense path dependent: the exact value to an agent of those other branches—of the chance benefit or the risk harm to which she has been exposed—is not strictly just a measure of their relation to the outcomes at those branches. And since in our model, although people can be irrational about the degree to which they value their winnings or losses, there is a fact of the matter as to the benefit for the agent of the chance of winning and a fact of the matter as to the harm for the agent of the chance of losing. The degree of benefit or loss will certainly be parasitic on other things the agent desires—the value she places on certain outcomes. But whatever the value, we can define the possibility of benefit on another branch of the plan as itself a benefit, and the risk of harm on another branch as itself a harm.

I have done my best to defend the related ideas of chance benefit and risk harm, though, as I have suggested, they must be defended in asymmetric form. While this may seem to cast doubt on the general approach, there is some intuitive support for the asymmetrical treatment of benefit and harm. Consider again the case of the lottery ticket. Suppose you have paid \$1 for a lottery ticket in a million dollar lottery, and I then steal the ticket. If the ticket is not the winning one, presumably I must compensate you for the price of the ticket, namely \$1. That seems a strong intuition in favor of the idea of chance benefit, for in its absence, we must conclude that the person who steals what turns out to be a losing lottery ticket owes no compensation. What, however, do I owe you in the case of a winning ticket? The possibilities are: (1) a million dollars, (2) a million dollars plus the price of the ticket, or (3) only the price of the ticket. I take it that (3) can be summarily eliminated, as drastic undercompensation, so that our only real possibilities are between (1) and (2). Now it seems to me it would be odd to say I owe you a million dollars plus the price of the ticket, or \$1. This would be double counting, given that the price of the ticket is

only payment for the chance of the million dollars. If you gain a million dollars on the basis of the \$1 ticket, your ticket price is returned to you in the value of the winnings.

Moreover, if I must pay you a million and one dollars by way of compensation, you get a windfall, since if I had not stolen a ticket that turned out to be the winning ticket, you would only get a million dollars, not a million and one. So the only argument we have for saying I owe you a million and one is by inference from the case in which I steal the losing ticket. But I have tried to make intuitive that we can treat chance benefit and risk harm asymmetrically with respect to winning and losing.

A second problem, however, is not solved by adopting the asymmetric approach, namely, the problem of infinite regress. Suppose we count the chance benefit in the losing payoffs to an agent from a plan, along with the outcome benefit. Then it looks as though there must be a gamble that it would be rational for an agent to accept that would give her that combined payoff of chance benefit plus outcome benefit. Let us say the payoff from this gamble is $p = c + o$ (payoff = chance benefit + outcome benefit). There is, however, a benefit to the agent from being able to enter such a gamble. Let us call this new chance benefit c' . From that it follows that there is a p' , such that $p' = c' + c + o$. There is a benefit to the agent from being able to enter the gamble that would give her *that* payoff, whose payoff is $c'' + c' + o$, and so forth. Now one might be tempted to think this is not the problem it appears to be. For there can be an entry fee to a poker game, and by the same token there could be an entry fee to gain membership in a group from which the participants in a poker game might be chosen, and so on. There is no problem with the idea of chances over chances. But here the regress is more pernicious. For it means that the payoffs from the original gamble are not well defined. The payoffs from a risky plan, say, on the losing branch would be the outcome at that branch, plus an infinite series of ever-increasing chance benefits. In this case, the payoff from any gamble would paradoxically be infinitely high if the agent lost, and infinitely low if the agent won!

While this is a serious objection, it does not compel the rejection of the ideas of chance benefit and risk harm. Instead, it requires that we reject the idea that chance benefit is something the agent can aim at as part of the benefit she receives from a risky plan, and that risk harm is something in and of itself that an agent seeks to avoid. That is, while the chance she had at receiving an outcome benefit is itself a benefit she receives from a gamble, we cannot think of it as the kind of benefit an agent would pursue in its own right. Nor can we think of risk harm as a negative payoff the agent would structure her life to avoid. Strictly speaking, then, we should probably not think of chance benefit as the sort of thing that can be added to outcome benefit, since they are benefits of two very different kinds. It is not really part of the payoff from a gamble. If we could

add them, this would suggest that an agent would agree to trade off a bit of outcome benefit for some amount of chance benefit, even once the gamble has been completed. But no rational agent would do such a thing, and indeed, it is not clear the idea of trading off outcome benefit for chance benefit can be made coherent. Thus the numbers we have assigned to measure chance benefit in the illustrations of the chance benefit account offered above are only a heuristic for figuring out which gambles would be rational and which not.

This last point will undoubtedly leave the notions of chance benefit and risk harm somewhat elusive. If these are not part of the payoffs from a gamble, in what sense are they benefits or harms? We have no trouble saying that to give someone a lottery ticket is to benefit him before the lottery has been determined. The benefit in question, then, is simply the kind of benefit that the recipient of a lottery ticket has, but persisting over time. Once benefited with a chance of winning an outcome benefit, or harmed with a risk of suffering an outcome harm, we can continue to think of the person who has received the chance as benefited, and the person subjected to the risk as harmed, even once we know that the chance or risk has failed to eventuate. In general, the thought is that we are better off for having opportunities for gain in our lives—that the opportunities are themselves beneficial—even if those opportunities do not eventuate in favorable outcomes. Similarly, we are worse off for being exposed to the possibility of loss, even if the loss does not eventuate.

VII. CONCLUSION

It is important to be clear about the scope of my conclusion. First and foremost, I take myself to have shown that the pragmatic theory does not deal effectively with plans involving an element of risk and that despite the apparently compelling distinction between assurances and threats Gauthier draws in “Assure and Threaten,” no such structural distinction obtains. From the standpoint of Gauthier’s theory, the distinction between plans involving an element of risk and plans whose benefits are certain to accrue is significant. This is not surprising: any theory of practical rationality that requires rational actions to be identified by their relation to optimal plans, as opposed to plans with optimal expected utility, will have difficulty explaining the rationality of gambles.

Second, I have argued that on an objective approach to probabilities, articulated in terms of relative frequencies, there is a way of conceiving of opportunities to benefit as positive additions to an agent’s baseline welfare, as opposed to merely epistemic (i.e., perceived) additions to welfare. In some probabilistic cases, I have argued, the objective, *ex ante* chance of benefit augments the level of outcome welfare that results from a losing gamble.

Third, I have suggested that there may be cases in which the chance benefit is sufficiently great, and the level of welfare loss relative to one's original baseline is sufficiently low, that the chance adjusted welfare on the losing branch overwhelms the agent's losses from the decrease in welfare. I have tried to provide a model of how this might be and to suggest analogies, such as familiar cases of winning a chance to win some outcome benefit (examples might include winning a lottery ticket or admission to a poker game) that could make sense of this claim about the value of chance benefit. If such cases exist, some risky assurances would be rational to offer, just as some deterrent threats are rational to make. And the rationality of such gambles would not depend on the expected utility the agent gets from the chance of standard outcome benefit.

Finally, however, we should be clear that even if there are cases of rational risky assurances where the chance benefit makes the plan optimal even in the absence of outcome benefit, their existence does not necessarily solve the problem of risk that Gauthier identifies in "Assure and Threaten." The central difficulty is that chance benefit adds only incrementally to the value of a losing gamble. Thus it remains the case that most losing gambles are not eligible for rational adoption on Gauthier's version of the pragmatic theory. Other pragmatic accounts, however, will either have the same difficulty or will suffer from the defect that they fail to satisfy the Pragmatic Deliberative Principle. Such is the case with Gauthier's *Morals by Agreement* approach and other accounts that make use of nondeliberative mechanisms for plan execution.

It bears noting that pragmatic theories are not truly in worse shape than expected utility theories in this regard. For any case with the structure we have been considering—where the suboptimal action called for by the plan is not instrumentally necessary for achieving the benefits from the plan—will also be one in which it is irrational to follow through on the plan on standard expected utility theories. Proponents of the standard approach to rationality are admittedly not bothered by this fact. They see rationality as process constrained in such cases and do not determine the plausibility of the outcome by the benefits achieved under the plan. Given that pragmatists all share the intuition that this result counts heavily against expected utility theory, they are hardly in worse shape under a deliberative version of the pragmatic approach, combined with the chance benefit and risk harm theses.