Deterrence and power

GLENN H. SNYDER

Princeton University

Anyone who comes to the study of military affairs from the field of international relations is likely to be struck by the discontinuity between the prevailing theoretical concepts in the two fields. One notes, for example, that, although a good deal of theoretical effort has been devoted to the idea of deterrence in the context of the problem of national security, it is still not clear where this notion stands in relation to the older concept of power, which also has been subjected to considerable analysis. One also wonders how the ancient idea of “balance of power” should be modified to take account of the idea of nuclear “parity” or “balance of terror.”¹ The literature on national security abounds with such concepts as “active” and “passive” deterrence and defense, “counterforce” and “retaliatory” strategies, and alliance “burden-sharing,” all of which have yet to be integrated into international relations theory. While this paper is chiefly concerned with probing into the nature and meaning of deterrence, I hope in passing to show the connection between this concept and the more general idea of power.

Conceived broadly, deterrence appears to be a species of “political power.” Defining political power generally as the capacity to induce others to do things or not to do things which they would not otherwise do or refrain from doing, deterrence is simply its negative aspect. It is the power to dissuade another party from doing something which one believes to be against one’s own interests, achieved by the threat of applying some sanction.

Deterrence does not have to depend on a threat and capacity to impose punishment. It may also be achieved by having the capability to deny the other party any gains from the move which is to be deterred. Thus we may speak of “deterrence by denial” as well as “deterrence by punishment.” In military affairs deterrence by denial is accomplished by having military forces which can block the enemy’s military forces from making territorial gains. Deterrence by punishment grants him the gain but deters by posing the prospect of war costs greater than the value of the gain.

Deterrence (like political power) does not have to depend on military force. We might speak of deterrence by the threat of trade restrictions, for example. When non-military means are used, deterrence may be accomplished by the promise of rewards as well as by the threat of deprivation. Thus the promise of economic aid might “deter” a country from military action (or any action) contrary to one’s own interests. Or we might speak of the “deterrence” of allies and neutrals as well as potential enemies, as Italy, for example, was “deterred” from fighting on the side of

¹ A remarkable pioneering effort to assess the impact of the new weapons on the classic balance of power was Arthur Lee Burns’s article, “From Balance to Deterrence” (2).
the Dual Alliance in World War I by the promise of substantial territorial gains. Both deterrent capability and political power stem from the control which one has over another party’s total value “inventory,” whether the control takes the form of a capacity to increase or to decrease this inventory.

Robert A. Dahl (3) has written that power consists of four basic components: base, means, amount, and scope. The power base is the material or attribute which provides the capacity to affect the value positions of others, for example, military force, economic assets, etc. The means is the method by which the power base is brought to bear, for example, by threat, ultimatum, or force demonstration. The amount of power is the degree of influence over potential actions. The scope is the range of potential actions by the other party which can be influenced by the threat or promise of applying the base. Degree of influence is expressed most precisely as a probability figure indicating the chances that the other party will do (or refrain from doing, as the case may be) the things which the power-wielder specifies in his threat or promise.

These categories can also be applied to the phenomenon of deterrence. Considering, for example, deterrence by “massive retaliation,” the base is the capacity to inflict devastating punishment, the means is the threat of retaliation, the scope refers to the various forms of aggression the probability of which may be reduced by the threat, and the amount is the reduction in probability of each of these moves which results from the threat.

I would add two further components to the four suggested by Dahl: the object values and the credibility of a threat or promise. Object values are the values of the other party, which are subject to being decreased or increased by the actual carrying-out of the threat or promise. Object values are theoretically distinct from both power base and scope values. The scope of the power-wielder’s power follows from the threat of deprivation of, or the promise of additions to, the other party’s object values by application of a power base. Thus deterrence of an aggressive act (scope) may be effected by the threat of applying nuclear punishment (power base) to the aggressor’s cities and population (object values).

The juxtaposition of deterrence and power helps one to recognize the importance of a sixth component, credibility, defined as the perception by the threatened party of the degree of probability that the power-wielder will actually carry out the threat if its terms are not complied with or will keep a promise if its conditions are met. The idea of credibility, so prominent in the literature on deterrence, really has application to all forms of political power. In domestic politics the person or group against whom a threat is directed may harbor doubts about whether the threatener “really means” it, and these doubts will affect the amount of power actually available to the threatener. A President may threaten to withhold patronage from a particular senator, for example, but, if the senator has some power of retaliation (e.g., by failing to support other aspects of the President’s program) or if the withholding of patronage might adversely affect the President’s political support in the senator’s state, the President’s threat may not be very credible to the senator, and his voting on the immediate issue may not be affected.

In short, the concept of power must take some account not only of the power-wielder’s control over the value inventory of the recipient of the threat but also of the latter’s capacity to affect the values of the power-wielder and, generally, of all the possible adverse consequences to the power-wielder from carrying out his threat. Political power, like deterrence, is a two-way street; A’s power over B depends on B’s power over A.
DETERRENCE AND POWER

little or no power over B, even if A possesses a "base" for inflicting deprivations on B, if B can inflict punishments of similar weight against A and can be expected to do so in retaliation with a high degree of credibility. Of course, if the credibility of B's retaliation is low, A may still have an edge in the "balance of power" vis-à-vis B, but the amount of his political power over B (defined as the probability that B will behave in the manner desired by A) must be discounted, in some sense, by the loss of credibility occasioned by B's recognition of A's recognition that B might retaliate and also by the size of the value loss which both know B can inflict in retaliation.

All that I have been saying is that the source of political power resides in two distinct elements: capabilities (or the capacity to affect object values by application of a power base) and the opponent's perception of the intent to use these capabilities if one's demands are not met (4). In the field of international relations, both in scholarly analysis and in actual practice, the intentions factor has often been given less attention than it deserved. What was always unreal, for example, about the argument for building up "situations of strength" in Europe to make the Russians more amenable to negotiations was that it seemed to be assumed by its proponents that "strength" by itself was sufficient to extract concessions from the enemy. It was overlooked that the strength would have to be accompanied by a willingness to use it, and, when one stopped to think about it, it seemed doubtful that the West was, or would be, so willing. Similarly, the United States was never able to extract much political power from the atomic bomb during the heyday of its monopoly because it was clear to all that there was no intent to use it except in response to the most serious aggression.

The advent of "nuclear parity," defined as a situation in which the recipient of an all-out nuclear strike would have the means to deliver completely unacceptable retaliation, has vastly increased the importance of the intentions factor in international power relations. In earlier ages the important calculation for the incipient aggressor concerned the balance of capabilities between himself and his opponent. Would he, or would he not, be able to win? It was rather certain that the defender, and, with somewhat less certainty, his allies, would respond with whatever force might be available. But now that the destructive power of military force has far outrun almost all conceivable interests or political goals, the response, and the severity of the response if it comes, is far less certain. What is certain for both sides is that the other has the capability, if he cares to use it, to exact costs which render insignificant any potential territorial changes. The critical calculation now concerns the balance of intentions. Each side must ask: What portion of his available destructive power will my opponent use in response to this or that move of mine? How willing is he to take steps which may increase the danger of unacceptable destruction for both of us? How does his willingness to run risks compare with mine? Is the opponent likely to act irrationally? How does he think we compare in "nerve," in "will power," in propensities to irrationality?

I have spoken of "political power" to preserve a useful distinction between the power to persuade which results from the threat or promise to cause deprivations or additions to object values and the physical power to affect object values. To avoid overburdening the word "power," it seems useful to label as "capability" the direct results to be expected from actual use of a power base, that is, the implementation of a threat or promise. Thus in international affairs we distinguish between "military capability" in war and the "political power" in advance of war which flows from the possession of, or threat to use, military force. But to think of deterrence (or
political power) as something which operates only in advance of war is not entirely valid. It would be only if war were an absolutely unrestricted donnybrook, continued until one side is physically unable to continue. If war is like this, it is obvious that deterrence stops once the war starts. But, in the past at least, wars usually have not been like this. Typically, their outcome has been determined by a combination of political power and military capability. In other words, one side accepts defeat as the result of a combination of “raw” physical coercion, and the expectation or realization that continued fighting can only generate additional costs without hope of compensating gains, this expectation being largely the consequence of previous application of raw force by the dominant side. Thus deterrence operates in war as well as before war. This can be seen most clearly in proposals for a strategy of limited strategic retaliation, which would have as its primary purpose convincing the enemy that continuing the war would be unprofitable for him because it would cause further nuclear strikes against his homeland. The initial strikes are in effect threats of further strikes to come, designed to “deter” the enemy from further fighting. One should also note that, in limited warfare in the nuclear context, the nuclear weapons held in reserve by either side constitute a deterrent against the other side’s expanding the intensity and destructiveness of its war effort.

But to identify deterrence with negative political power is still too narrow a definition of deterrence. Deterrence is a function of the total cost-gain expectations of the party to be deterred, and these may be affected by factors other than the apparent capabilities and intentions of the deterrer. For example, the incipient aggressor may be inhibited by his own conscience or, more likely, by a prospect of losing moral standing, and hence political standing, with uncommitted countries. Or, in the specific case of the Soviet Union, he may fear that war will encourage unrest in, and possibly dissolution of, his satellite empire and perhaps disaffection among his own population. He may anticipate that his aggression would bring about a tighter welding of the Western alliance or stimulate a degree of mobilization in the West which would either reduce his own security or greatly increase his own cost of maintaining his position in the arms race. It is also worth noting that the benchmark or starting point for the aggressor’s calculation of costs and gains from military action is not his existing value inventory but the extent to which he expects that inventory to be changed if he refrains from initiating military action. Hence the common observation that the Russians are unlikely to undertake overt military aggression because their chances are so good for making gains by “indirect” peaceful means. Conceivably, the Soviets might attack the United States, even though they foresaw greater costs than gains, if the alternative of not attacking seemed to carry within it a strong possibility that the United States would strike them first, in which case their own costs would be enormously higher. In a very abstract nutshell, the potential aggressor presumably is deterred from a military move not simply when his expected cost exceeds his expected gain but when the net gain is less or the net cost is more than he expects when he refrains from the move. But this formulation must be qualified by the simple fact of inertia: deliberately to shift from a condition of peace to a condition of war is an extremely momentous decision, involving incalculable consequences, and a government is not likely to do so unless it foresees a very large advantage in making this move. The great importance of uncertainty in this context will be discussed below.

These rather discursive observations have been by way of establishing some connection
between deterrence and power. Deterrence, it seems, is a form of power relation—the power to dissuade—but it is also broader than power in that it can result in part from factors which are not part of the deterrer's power base. The theoretical utility of the term is enhanced, however, if deterrence is defined more narrowly as the discouragement of military aggression by the threat (implicit or explicit) of applying military force in response to the aggression. This will be the meaning of the term in the remainder of this paper. I shall turn now to a somewhat more rigorous analysis of the idea of deterrence.

The Logic of Deterrence

Deterrence, like all political phenomena, is not likely to operate according to strict logic. However, it does have a logical core which is useful as a basis for analysis.

The object of military deterrence is to reduce the probability of enemy military moves inimical to one's self. If we postulate two states, an "aggressor" (meaning potential aggressor) and a "detterer," the probability of any particular attack by the aggressor is the resultant of four factors which exist in his "mind." All four taken together might be termed the aggressor's "risk calculus." They are (1) his valuation of the territorial objective; (2) the cost which he expects to suffer in an attack on the objective, as the result of various possible responses by the deterrer; (3) the probability of various responses, including "no response"; and (4) the probability of winning the objective with each possible response. These are the basic factors the aggressor must weigh in determining, on balance, whether a venture is likely to result in a net gain or a net cost.

The third factor in the enemy's calculus represents the "credibility" of various possible responses by the deterrer. But credibility is only one factor; it should not be equated with the deterrent effectiveness of a possible or threatened response, which is a function of all four factors, that is, the net cost or gain which a response promises, discounted by the probability (credibility) of its being applied. An available response which is very low in credibility might be sufficient to deter if it poses a very severe sanction (i.e., "massive retaliation") or if the enemy's prospective territorial gain carries relatively little value for him. Or a threatened response that carries a rather high credibility but poses only moderate costs for the aggressor (e.g., a conventional response) may not deter if the aggressor places a high value on his objective and anticipates a good chance of successful conquest.

The "credibility" factor deserves special attention, however, because it is in terms of this factor that the "risk calculus" of the aggressor "interlocks" with that of the deterrer. The deterrer's risk calculus is similar to that of the aggressor. If the deterrer is rational, his response to aggression will be determined (within the limits, of course, of the military forces he disposes) largely by four factors: (1) his valuation of the territorial objective and of the other intangible gains (e.g., moral satisfaction) which he associates with a given response; (2) the estimated costs of fighting; (3) the probability of successfully holding the territorial objective; and (4) the change in the probability of future enemy attacks on other objectives which would follow from various responses. Variations on, and marginal additions to, these factors may be imagined, but these four are the essential ones. The deterrer will select the response which minimizes his expectation of cost or maximizes his expectation of gain.

The aggressor, of course, is not omniscient with respect to the deterrer's estimates of cost and gain. Even the deterrer will be unable to predict in advance of the attack how he will visualize his cost-gain prospects once the aggression is under way. (Witness the
United States response to the North Korean attack in 1950, which was motivated by values which apparently did not become clear to the decision-makers until the actual crisis was upon them.) Nor can the aggressor be sure the deterrer will act rationally according to his own cost-gain predictions. Because of these uncertainties the aggressor’s estimate of credibility cannot be absolute. More than one response will be possible, and the best the aggressor can do is attempt to guess how the deterrer will visualize his gains and losses consequent upon each response and from this guess arrive at a judgment about the likelihood or probability of each possible response.

The deterrer, in his turn, evaluates the effectiveness of his deterrent posture by attempting to guess the values of the four factors in the aggressor’s risk calculus. In estimating the credibility factor, he attempts to guess how the aggressor is estimating the factors in his (the deterrer’s) calculus. He arrives at some judgment as to whether the aggressor is likely to expect a net cost or net gain from the aggressive move, and, based on this judgment and his degree of confidence in it, he determines the probability of aggression. Happily, the spiral of “guesses about the other’s guesses” seems to stop here. In other words, the aggressor’s decision whether or not to attack is not in turn affected by his image of the deterrer’s estimate of the likelihood of attack. He knows that, once the attack is launched, the deterrer will select the response which promises him the least cost or greatest gain—at that point the deterrer’s previous calculations about “deterrence” of that attack become irrelevant.

It is possible to express these relationships mathematically. The figures which follow are based on the assumptions that both sides are able to translate all relevant values into a “common denominator” utility, that they can and do estimate probabilities for each other’s moves, and that they act according to the principle of “mathematical expectation.”2 The latter states that the “expected value” of any decision or act is the sum of the expected values for all possible outcomes, the expected value for each outcome being determined by multiplying its value to the decision-making unit times the probability that it will occur. To act “rationally,” according to this criterion, means simply to choose from among the available courses of action, the one which promises to maximize expected value (or minimize expected cost) over the long run. There are reasons—apart from the practical difficulty of assigning numerical values to the elements involved—why the mathematical expectation criterion is not entirely suitable as a guide to rationality in deterrence and national security policy. However, it is useful as a first approximation; the necessary qualifications—having to do largely with the problem of uncertainty and the disutility of risking large losses—will be made presently.

Imagine a world of four states: A, B, C, and D. By a happy alphabetical coincidence, we

2 The numerical illustrations are intended simply to set out as starkly as possible the essential logic of deterrence; there is no intent to light a torch for the “quantifiability” of the factors involved, which are, of course, highly intangible, unpredictable, unmeasurable, and incommensurable except in an intuitive way. It is worth keeping in mind, however, that decision-makers do have to predict, to measure, and, in some sense, to make incommensurable factors commensurate if they are to reach wise decisions. Although, in practice, the factors cannot be given precise numbers, it is legitimate, for theoretical purposes, to pretend that they can be in order to clarify the logic or method by which they should be weighed and compared. The logic is just as applicable to imprecise quantities as to precise ones; to express it in mathematical terms can provide a useful check on intuitive “judgment” and may bring to light factors and relationships which judgment would miss.
can say that A is the "aggressor" and D the "deterrer." (Those who prefer not to think in abstractions may think of A as the Soviet Union, B as western Europe, C as "non-Communist Asia," and D as the United States.) Both A and D are thermonuclear powers. They are in a condition of "nuclear stalemate," that is, neither, by striking first at the other, can prevent the other from striking back with an effect outweighing any possible gains. We assume, therefore, that surprise attack on D is not a rational move for A, since D is practically certain to retaliate. In the first illustration (Figs. 1 and 2), which we might call the "massive retaliation" model, A has substantial conventional ground forces, but D has none. Both B and C are the objects of contention between A and D, and they are allied to D. Neither B nor C has nuclear or conventional forces. We assume that both A and D are rational but that they are not certain of each other's rationality.

To simplify, we postulate that A has only two possible moves: to attack B with full conventional strength or not to attack at all. Similarly, D has only two available responses: to "massively retaliate" or to do nothing, that is, to acquiesce in the loss of its ally. The cost of all-out war for both sides is 100, and the value of B to both A and D is 20. We assume, for simplicity, that an all-out response will preserve the independence of B.

Figure 1 represents A's cost-gain estimates ("payoffs") and the probabilities he associates with each of D's possible responses to an attack. Figure 2 represents D's payoffs and his estimate of the probability of attack. For convenience, we have made their valuations of gains and losses symmetrical.

A estimates the probability of retaliation by D by attempting to guess D's payoffs. This probability, in theory, is determined by the size and direction of the "gap" between D's two payoffs, as estimated by A. If A estimates D's payoffs to be roughly equal, logi-

<table>
<thead>
<tr>
<th>D</th>
<th>Attack</th>
<th>Not Attack</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Retaliate</td>
<td>-100</td>
<td>-</td>
</tr>
<tr>
<td>Not Retaliate</td>
<td>+20</td>
<td>0</td>
</tr>
</tbody>
</table>

Expected values: +8 0

Fig. 1

D's Calculus

\[
\begin{array}{c|cc}
\text{D} & \text{Attack} & \text{Not Attack} \\
\hline
\text{A} & .60 & .40 \\
\text{Retaliate} & -100 & - \\
\text{Not Retaliate} & +20 & 0 \\
\end{array}
\]

\[
\text{FIG. 2}
\]

\[
.10(-100)+.90(20)=8.
\]

3 Strictly speaking, not only the absolute size of the gap but also the ratio between the size of the gap and the size of the payoffs would have to be considered.
If D were able to divine A's calculations, he would expect an attack with certainty. But on the best evidence available to him about A's payoffs and probability estimates (the latter depending on how he thinks A visualizes his own payoffs), the most he can say is that A's "expected value" is probably positive and small. This leads him to consider an attack more likely than not, but not much more. Therefore, he assigns a probability of .60 to the attack. This is D's measure of the "effectiveness" of his deterrent posture.

It is useful to note that the probability of retaliation at which A will be indifferent between attacking or not (at which his expected value for "attack" is zero) is approximately .17. From D's point of view, this is the "required credibility" of his threat, that is, the credibility necessary just to make it an effective deterrent. Required credibility depends on A's payoffs; credibility depends on A's image of D's payoffs. Obviously, D wishes at all times to keep "credibility" higher than "required credibility." If it is lower, as in the threshold, which involves changing A's payoffs, or by increasing credibility, which requires shifting his own payoffs or, more accurately, changing A's image of them. He can also try to increase A's doubts about his (D's) capacity to act rationally.

Let us consider first how D may affect A's payoffs and thus influence required credibility. Obviously, he can increase his nuclear striking power and thus increase the costs which he can inflict on A in retaliation. If his capabilities are already at the level where they can completely destroy A's society and productive economy, such a move would have little effect on A's calculations. But let us assume that they are not. Then an increase might conceivably raise A's all-out war costs to 200, as shown in Figure 3.

This change (assuming that A's capabilities do not change) makes D's nuclear deterrent effective by reducing required credibility to .09. Or, to put it another way, it makes A's expected value negative. Of course, if A matches D's increase, the actual credibility of D's threat of retaliation may also decline. But there is no inherent reason why, with similar increases in armament on both sides, these effects should exactly offset each other. They are more likely to do so if both parties are symmetrical in their estimates of the consequences of all-out war or, more exactly, if the aggressor believes such symmetry exists. But if A thinks D is less concerned than himself about increases in potential damage in all-out war—if the "marginal cost" of such damage is less for D than for A, perhaps because D has a much better civil defense program—a rise in the nuclear striking power on both sides would be likely, on balance, to increase D's deterrent power. Required credibility, in other words, would fall more than credibility. Such a relationship is illustrated Figure 1, his opponent is likely not to be deterred. He can attempt to bring credibility over the required threshold either by lowering the threshold, which involves changing A's payoffs, or by increasing credibility, which requires shifting his own payoffs or, more accurately, changing A's image of them.
in Figure 4. The intersection of the lines represents the point at which D's deterrent power just begins to be effective; reciprocal increases in striking power beyond this point will increase D's deterrent effectiveness, since credibility exceeds the credibility requirement by a greater and greater margin. Of

This change in D's capabilities forces A to consider three possible initial responses: massive retaliation, tactical nuclear warfare, and a conventional surface response. "No response" is ruled out because D's ground forces are practically certain to fight if they are attacked. A expects net costs of 20 in the event of a tactical nuclear response, which he calculates by considering two possible outcomes and their probabilities, should D respond initially with tactical nuclear weapons: a tactical nuclear war which stays at that level and an initial tactical nuclear response which spirals to all-out war.⁶ The fact that D has created an alternative to an all-out war leads A to reduce the probability of an initial all-out response to .05. A ascribes a much higher probability to a conventional response than to a tactical nuclear one because he believes

---

⁶ If A's cost of all-out war is 100, if his estimated net cost of tactical nuclear war is 20, and if he thinks that there is a .25 chance that a war which starts at the tactical nuclear level will eventually become all-out, the calculation is .25(−100) plus .75(−20) plus 20 equals −20.
D fears the consequences of tactical nuclear warfare. Figure 6, which shows D's payoffs, indicates that A is correct in this assumption. A estimates a net gain of 12 if D responds conventionally. This figure also is calculated from a "cluster" of "sub-outcomes" and their probabilities, the salient ones being that the conventional response will stay limited, that it will grow to tactical nuclear war and stay at that level, or that it will eventually spiral to all-out war. A's entire calculation yields him an expected value of about minus 3, which, of course, deters him.

The chief point to be noted about this last example is that over-all deterrent effectiveness can be increased by the provision of ground forces which, although they cannot hold the objective and they depreciate the credibility of an immediate massive retaliation, do insure some violent response and pose risks of still greater eventual violence—risks which the aggressor is unable to accept.

By holding constant the probability of any one of the three initial responses, a "required credibility" can be calculated for either one of the other two. For example, if the probability of massive retaliation is held at .05, the probability of a tactical nuclear defense must be at least as high as .20 for effective deterrence. Or, if the probability of the tactical nuclear response is held at .30, the probability of immediate massive retaliation must be at least .03.

If we introduce the assumption that A is uncertain whether he can "win" (i.e., conquer B) in a tactical nuclear war, he must discount the value of his possible territorial gain (20) by the probability that he will be unsuccessful. This raises the over-all "expected cost" of the tactical nuclear response for A because he has less potential gain to set off against his estimate of potential loss and risk.

We might also note that A's valuation of the objective affects required credibility, as D's valuation affects credibility. For example, if the object of contention is C rather than B, and C is valued at only 10 by both sides, required credibility in the massive-retaliation model (Figs. 1 and 2) is only .09, rather than .17, as it is when B is the object. But, since A recognizes that C carries also a lower value for D, he will attach less credibility to the massive-retaliation threat. If the comparative valuations of B and C are symmetrical for both sides, there is a prima facie case for believing that the relation between required and actual credibility, and hence the deterrent effectiveness of the threat, is roughly the same for both objects. This offends common sense, which seems to indicate that such very violent threats as that of massive retaliation have lower effectiveness as well as lower credibility when the prize is small than when it is large. This intuitive appreciation rests on the assumption—a plausible one—that the potential aggressor ignores very small probabilities; that probabilities of retaliation as low as, say, one chance out of twenty, are considered equivalent to zero in his calculus.

This paper is chiefly concerned with the deterrence as an attribute of military power, but at least passing reference should be made to the deterrent usages of diplomatic declarations and threats. Threats by the deterrer may serve to impress upon the aggressor that a certain objective is valued highly by the deterrer. Threats may also change the deterrer’s payoffs and consequently enhance the credibility of the threatened action. They do so by increasing the cost of not responding in the threatened way, by implicating additional values beyond the bare value of the territorial objective—values which would be lost if the threat is not carried out. For example, in our

7 For a lucid discussion of threats, commitments and bargaining see Thomas C. Schelling, “The Strategy of Conflict: Prospectus for a Re-orientation of Game Theory” (10).
massive-retaliation model (Figs. 1 and 2), if D were to pledge his honor and prestige in an unambiguous threat to retaliate, the cost of failure to retaliate might be raised from 20 to 30 or 40. Retaliation would still be irrational for D, but A would be less certain of this than he was before. Fearing that, in making the threat, D might have increased his cost of not retaliating above his cost of retaliating, he would assign a somewhat greater probability than .10 to the retaliation outcome.

In addition to the intrinsic losses in failing to carry out a threat, there may also be political losses in the form of reduced credibility of other threats and a reduced capacity to attract allies. In making a threat, in other words, the deterrer places in hostage not only his own honor and moral self-respect but also certain aspects of his future deterrent power. In doing so, he makes it rationally more difficult to fail to make the response, and the aggressor, recognizing this, expects a response with greater probability.

The credibility of an irrational response may be increased if the deterrer can appear to commit himself to this response by some device which removes or reduces his freedom of choice. Such "automation" is itself rational—even though, paradoxically, the response is not—if the aggressor can be expected to believe with high confidence that the commitment is irrevocable and is thereby deterred. Automation differs from threat-making in that it does not change the underlying payoffs but rather inhibits the choice of all responses except the one which is being relied on for deterrence. Complete mechanical automation is probably impossible in military affairs, but it can be approached in various ways. The use of ground troops as a "trip-wire" for nuclear retaliation is a kind of automating device. A military commander may be given advance authority to order retaliation as soon as he sees that an attack is under way. Or military organization and planning may become so oriented around a particular response that the inertia in its favor may be very difficult to overcome in a sudden crisis. The integration of tactical atomic weapons into NATO forces, and the plans to use them in case of major aggression against western Europe, is a clear case in point. The commitment to massive retaliation may also be subject in some degree to such administrative automation.

The enemy does not have to believe that automation is complete—that freedom of choice has been entirely relinquished—in order for deterrent effects to accrue. Even if the deterrer can only commit himself partially to an irrational response, that is, restrict his freedom of choice but not eliminate it, doubts may arise in the aggressor's mind as to whether the deterrer has retained enough flexibility to be able to overcome the built-in bias in favor of this response. One might say that the United States commitment to resist aggression in the Middle East—contained in the "Eisenhower Doctrine"—is at least partially automatic by virtue of the doctrine's provision that United States intervention will take place at the request of the country attacked.

The deterrer may increase the credibility of a seemingly irrational response by creating the general impression that he is prone to act irrationally. It is certainly paradoxical, even bizarre, to say that a certain amount of "irrationality" can be an aspect of "rationality." Yet this follows logically if deterrence is to depend on a threat which would be madness to carry out.

Leaving this paradox aside for a moment, "rationality" may be defined as choosing to act in the manner which gives best promise of maximizing one's value position, on the basis of a sober calculation of potential gains and losses and probabilities of enemy actions. This definition is broad enough to allow the
inclusion of such "emotional" values as honor, prestige, and revenge as legitimate ends of policy. It may be perfectly rational, in other words, to be willing to accept some costs solely to satisfy such emotions; but, if such emotions inhibit a clear-eyed view of the consequences of an act, they may lead to irrational behavior. It would be irrational, however, to satisfy a momentary passion if sober judgment would reveal that the satisfaction obtained would not in the long run be worth the cost.

Irrationality may take the form either of failing to act in accordance with one's best estimate of costs, gains, and probabilities or of faulty calculation of these factors in the light of the evidence available. Irrationality of the latter kind may stem from such sources as commitment to a dogma or theory which is inapplicable to the situation or which shuts out relevant data; education, training, and experience which prevents attainment of the "whole view"; or limited or distorted perspectives resulting from bureaucratic parochialism.

Irrationality may be in the direction either of recklessness or of timidity; naturally, for deterrence, one wishes to project an image of excessive boldness rather than excessive caution. But is it possible for a government to appear irrational and be rational? This would mean appearing to be willing to act contrary to one's payoffs, or appearing to have miscalculated them, when in fact one has calculated, and intends to act, rationally. It seems unlikely that a democratic government could credibly practice this kind of sophisticated deceit, but an authoritarian one might. In the United States it is difficult enough to achieve consensus about what is rational; to attempt to go further and practice "calculated lunacy" while secretly intending to act sanely seems infeasible. And, even if it were feasible, there would be a large chance of being found out in the pretense. It seems, therefore, that the only way a democratic government can appear irrational is actually to be irrational. This leads to some interesting thoughts. For example, even if I, as an individual, believe it would be irrational for the United States deliberately to initiate nuclear war in any circumstances, I may not be inconsistent if I also believe that it is rational for President Eisenhower not only to threaten but actually to intend to undertake such a war in defense of West Berlin. I may not wish to argue against this intent if I think that the deterrent effect is great enough. Or I may think it would be irrational for the United States to retaliate all-out against Soviet cities after a surprise attack which reduced SAC by three-fourths and left the Soviet Union with an enormous preponderance in striking power, but for the sake of deterrence I would not want the government to think so.

Of course, the deterrer must consider not only the usages of irrationality in his own policy but also the degree of irrationality of the potential aggressor. The latter's irrationality may affect his appreciation of the deterrer's potential irrationality. For example, if the Soviets really believe their ideological tenet that the "capitalist" elites are cool, calculating, and highly rational (7, pp. 18 and 23), any image of irrationality which we project, whether contrived or not, may be fruitless.

Some Qualifications

The "real world" is necessarily much more complex than any theoretical model which purports to describe or explain it. One might say that the chief purpose of theory is to abstract out the non-essential variables so that the interaction of the essential ones can be perceived in fullest clarity. I have attempted to present the logic of deterrence in its barest skeletal form; the skeleton now requires "fleshing-out" by way of certain qualifications.

The first two qualifications have to do with
the basic principle of "mathematical expectation." This principle is an approximately accurate and valid guide to decision-making only when the values or "stakes" with which it deals are small relative to the total value inventories of the decision-making units and when there is at least some evidence on which to base probability estimates. Obviously, the first condition does not hold in international relations in the nuclear age, when certain outcomes may involve catastrophic losses which cannot be recouped. A rational individual will be reluctant to participate in a game of chance in which he may lose his entire capital, even if his mathematical expectation is one of substantial net gain. How much more compelling is this consideration when the worst possible outcome is not just the loss of a fortune but the loss of millions of lives and the utter collapse of a society! Thermo-nuclear war simply does not fit well into the imagery of the gambling table. This does not completely invalidate the use of probabilities and "expected value" reasoning, but it forces some rather drastic modifications of the results of such reasoning.

Clearly, it would be going too far to say that a potential aggressor (assuming he is rational) will be deterred from all moves which involve even the slightest risk of all-out war. Intuitively, one feels that probabilities below a certain level tend to be disregarded. For example, if the Soviet leaders were to see only one chance in twenty that the United States would invoke massive retaliation against a large-scale conventional attack in Europe, they might very likely consider this possibility too insignificant to worry about—in effect, act as if the probability were zero. Or, to put it another way, probabilities below this level might be considered "normal risks" which should not influence behavior, much as one does not stop to consider the small statistical probability of disastrous accident when he takes his family out for a Sunday drive.

Above this threshold the probability of retaliation would have some operational meaning for the aggressor. When it does take on meaning (i.e., when the aggressor does consider it "high enough to worry about"), he is likely to subtract some indeterminate amount from the "expected value" of his move to compensate for the disutility of risking the heavy losses of all-out war. This consideration tends to reduce the "required credibility" of the deterrer's threat. The amount that is added presumably will have something to do with the aggressor's attitude toward gambling. Some individuals may be willing to risk their entire capital when the odds favor them only slightly. Others require a much higher expectation of gain. Presumably, similar differences in gambling propensity exist among nations. We have been told that the Soviet ideology counsels conservatism in risk-taking (7, pp. 17 and 68). If this ideological tenet influences governmental decisions, the threat of severe retaliation (or the possibility of a limited engagement ballooning to all-out proportions) is likely to have a deterrent effect vis-à-vis the Soviets considerably greater than its "mathematical expectation" would suggest. The disutility of large risks will also tend to raise the deterrer's standard of necessary effectiveness for deterrent threats and postures which, if not heeded, would produce high costs for himself.

The second qualification is more serious. I have assumed that the aggressor will have sufficient evidence bearing on the deterrer's payoffs to be able to assign precise probabilities to the responses open to the deterrer. He does have some evidence, of course, chiefly the record of the deterrer's past reactions to aggression, the existing size, composition, and the deployment of his military establishment, and his policy declarations, including in the latter the expressions of articulate public opinion in the deterring country and its allies. Supplementing these sources, the aggressor could turn to what evidence is avail-
able concerning the "national character" or "psychology" of the deterrer as it pertains to foreign and military policy. For example, in judging the probability of American nuclear retaliation to an attack in western Europe, the Soviets would be wise to take note of certain American attitudes: our deep emotional involvement with western Europe and our strong sense of honor, which might lead us to fulfill a commitment at whatever cost.

After making use of all these sources of evidence, the potential aggressor would still have only a set of very general inferences about the deterrer's probable behavior—not certain predictions or even statements of precise probability. The hypothetical stimulus (i.e., the contemplated aggression) is likely to be unique in many important respects; therefore, there is not likely to be any similar situation in the deterrer's action record. Observation of the deterrer's preparedness of course will indicate fairly clearly what the deterrer can or cannot do. But, within the range of what is possible, preparedness is not a reliable indicator of intent, chiefly because of the various capabilities that will be available and because some forces may be designed to deter rather than to provide a capability for action. Nor are threats a reliable indication of intent, for they may be uttered for deterrent purposes only—that is, they may be bluffs, and, moreover, they are likely to be vague.

The aggressor faces uncertainty at two levels concerning the deterrer's "payoffs": first, in his estimate of the opponent's estimate of the consequences of his response and, second, in his estimate of the opponent's valuation of the consequences. The Soviets, for example, cannot know what consequences the United States leaders see arising from the initiation of tactical nuclear war in Europe. Do we believe that the first bombs dropped would influence the Soviets to call off the war in fear of suffering further costs? Or do we believe that tactical nuclear war in Europe will inevitably spiral into all-out war? Many other possible expectations lying between these two extremes might be imagined. But beyond that, even if the Soviets knew how the United States estimated the consequences, they could not know how we valued them. They might well doubt, even if we did foresee a high probability that a limited nuclear war would not stay limited, that we had fully realized ourselves how much we valued the things that would be lost to us in an all-out war.

All these considerations tend to raise the question whether it is valid at all to introduce "probabilities" into the analysis of deterrence. Everything depends, of course, on whether the enemy thinks in probability terms. If the Soviets simply come to a flat decision as to whether massive retaliation (or whatever response they are appraising) will occur or will not occur, then the analysis of deterrence would be considerably different and rather simpler than we have presented it. The problem for the deterrer then would be to make sure that his military posture and threats posed greater costs than gains for the aggressor and to make sure that his threat was believed. There would be no need to speak then either of "levels of credibility" or of "expected value." Threats would be either credible or not credible. Of course, the deterrer would still face the problem of estimating how much evidence supporting the threatened intent (or absence of contrary evidence) would be necessary to achieve credibility.

I am inclined to believe, however, that decision-makers, Russians included, do implicitly think in rough probabilities. The phrase "calculated risk" connotes more precision than is probably actually practiced, but it does suggest at least an awareness of the relevance of probabilities and even some sort of rough-and-ready expected value calculation. Of course, the probabilities cannot be "objective" ones, such as one calculates in estimating the chances that a certain face will
DETERRENCE AND POWER

turn up when a fair die is thrown or in projecting a "frequency distribution" from a large number of identical cases. But potential aggressor and deterrior do have some information bearing on each other's intent—enough to establish at least rough orders of subjective "likelihood." Such subjective feelings of likelihood may become more precise in the actual crucible of decision than their verbal statement would suggest. If the Soviets consider a massive response to ground attack in Europe to be "unlikely" in the abstract, they will be driven to make some judgment of "how unlikely" when they sit down around a table actually to decide whether to order such an attack. If the judgment then becomes "possible, but highly improbable" this will carry a meaning for the participants which is close enough to a mathematical statement of probability to allow them to make some rough judgment concerning the "expected cost" of the venture, that is, some combination of "likelihood" and "value" (or cost) of different possible outcomes. If pressed, the decision-makers probably would be able to choose the probability figure which seemed to them the most plausible, even if they were careful also to assign a wide margin of error to their choice. At least they would be able to specify a range of probabilities which seemed to be more plausible than any others and equally plausible. Such a statement might take the form, "less than 10 per cent but not more than 50 per cent."

If we assume that the aggressor will see enough evidence in the deterrior's behavior to justify the assignment of a range of "most plausible" probabilities to possible responses, or if his subjective feelings of likelihood or uncertainty (whether or not based on evidence) in effect take the form of such a range, the next question to ask is: What use does the aggressor make of this estimate? If he is to calculate an "expected cost" for his venture, he needs, for working purposes at least, a precise probability, not a range of plausible ones. Does he "split the difference" between the extremes of the range? Or does he, more conservatively, choose to base his decision on the most unfavorable probability within the range of the plausible? Or, even more conservatively, does he base it on the most unfavorable probability which is at least conceivable, though not as plausible as some others? The answer would seem to depend again on the risk-taking propensities of the aggressor.

Since it is the aggressor who takes the initiative, it is he who must bear most of the burden of weighing the uncertainties. In the nuclear age when the eventual outcome of even the smallest border skirmish might be utter devastation, the aggressor's uncertainty is an important deterring factor. This does not mean, however, that it is uniformly desirable to "keep the enemy guessing." To increase the aggressor's uncertainty is desirable with respect to such deterrent threats as massive retaliation which he is inclined to disbelieve in the first place. To increase his uncertainty forces him to attach a somewhat higher working probability to the outcome "all-out war" in order to cover himself against miscalculation. But of course it would be better to increase his certainty in the other direction. And, if the aggressor is thought to be already fairly certain that aggression will produce costs for him greater than his gains, it obviously does not pay to reduce his certainty.

As we said earlier, deterrence is a two-way proposition. Prior to his first move, and in the process of making it, the aggressor will attempt to inhibit undesirable responses, perhaps by proclaiming limited objectives or by threatening to raise the intensity of the war if these responses are made. After the initial act of aggression, the positions of "aggressor" and "deterrior" would be reversed in the illustrations earlier presented. It is then up to the original "deterrior" to estimate the probabilities, and consequent costs for himself, of the
various counterresponses available to the original aggressor and to make his initial response accordingly. In theory he will have made this calculation before the war in determining his military and threat posture; for both sides the prewar calculations of costs, gains, and probabilities must be "whole-war" estimates, as best these can be made. The calculus for the initial move for both sides is a "collapsed" estimate of the consequences for a whole series of moves and countermoves. But these estimates will change with the progress of the war as both sides gain new information about the other's intentions and capabilities and their own potential costs and gains. Objectives, valuations, and emotional commitments may change, so that military actions which seemed irrational before the war become rational during its progress—or vice versa.

We have dealt exclusively with the phenomenon of deterrence. But of course the deterrer, in choosing his optimum military and threat posture in advance of war, must estimate not only the effectiveness of that posture for deterrence but also the consequences for himself should deterrence fail. In short, he is interested in "defense" as well as "deterrence" and in being able to fight a war with minimum costs and losses as well as in reducing its probability; his "security" is a function of both of these elements. Capabilities and threats which produce a high level of deterrence may not yield a high degree of security because they promise very high costs should war occur. Thus a posture which is limited to a massive-retaliation capability and threat may reduce the probability of war quite low but at the same time carry the risk of very high cost in war if the threat is carried out or high losses in terms of territory, prestige, and capacity to make future threats credibly if it is not carried out. The "security" of the deterrer—more accurately, his degree of "insecurity"—with respect to any given challenge is the "expected cost" which results from multiplying the probability of the challenge times the estimated total cost (in casualties, territorial losses, prestige, etc.) of the least costly available response. In Figure 2 the deterrer's "insecurity" can be calculated as —12 for the contingency there considered, and his over-all insecurity is the sum of such figures for all possible enemy moves. The deterrer's goal is not to maximize deterrence, or even to minimize insecurity, but to achieve the lowest possible aggregate of "insecurity" and the expenditure of resources and values to reduce insecurity—as these two entities are weighed and made commensurable by some sort of intuitive judgment.

REFERENCES

2. BURNS, ARTHUR LEE. "From Balance to Deterrence," World Politics, IX (1957), 494–529.