

More calculations about deterrence¹

CLINTON F. FINK

Department of Sociology, The University of Michigan

A central problem in post-World War II discussions of defensive military policies has been that of "deterrence," defined as "the discouragement of military aggression by the threat (implicit or explicit) of applying military force in response to the aggression" (Snyder, 1960, p. 167). Thus defined, the notion of deterrence involves the notion of behavioral influence by communicative means; a deterrent threat is a conditional threat of punishment for an action, where both the specified action and the carrying out of the punishment are undesirable to the one who makes the threat. The purpose of the threat is to prevent the undesired action without actually fighting. Deterrence is problematic because deterrent threats are not always successful; many instances have been recorded in which such a threat has been followed by the very action it was designed to prevent. This is simply a special case of the more general statement that no communicative means for influencing human behavior is uniformly effective under all conditions.

¹ This article was written during my tenure as a USPHS Postdoctoral Fellow at the Mental Health Research Institute. I am especially grateful to Karl W. Deutsch and George Levinger for their extensive comments on an earlier draft.

The practical problem which arises in this situation is: how can a government make its deterrent threats more effective? The scientific question which must be answered in order to solve the problem is: what are the factors which determine the effectiveness of a deterrent threat?

Common sense has a ready answer: A threat cannot be effective unless it is believed by the recipient, and it will not be believed unless it is credible; therefore, to be effective, a threat must be credible. This conclusion is one of the basic assumptions in much of the literature on deterrence during the past decade. The problem of deterrence thus becomes the problem of credibility: to make a deterrent threat more effective, it must be made more credible. But this only pushes the inquiry one step back. We must now ask: how can a government make its deterrent threats more credible? And the prerequisite scientific question now becomes: what are the factors which determine the credibility of a deterrent threat?

Once again, common sense has an answer: a threat will be credible if there is reason to believe that the one who makes the threat is both *able* and *willing* to carry it out. This is similar to the conditions for credibility of any statement of intentions,

for example, the government's promise to exchange metal money for each piece of legal paper money. Thus, in order for a deterrent threat to be credible, it must be imbedded in a context of information concerning the threatener's ability and willingness to carry it out. These are the main factors cited in the literature which discusses the determinants of credibility (e.g., Schelling, 1960; Kaufmann, 1954; Snyder, 1960).

Most of the literature on deterrence has not been based on systematic empirical investigation, but has remained analytic, speculative, or normative. Recently, however, Russett (1963) presented the results of a comparative study of six historical cases of successful deterrence and eleven unsuccessful deterrence attempts, in which he explored the question of what makes a threat credible. The deterrence situations he examined were a special class defined as "all the cases during the last three decades where a major power 'attacker' overtly threatened a (minor power) pawn with military force, and where the (major power) defender either had given, prior to the crisis, some indication of an intent to protect the pawn or made a commitment in time to prevent the threatened attack" (p. 98). Thus, instead of ordinary self-defense, these cases involved the defense of a third party.

By comparing data on cases where the deterrent threat succeeded with data on cases where it failed, Russett arrived at a number of conclusions concerning factors which determine the credibility of such threats. In the present paper, I will argue that Russett's data do not compel assent to his conclusions, first because the observed dependent variable is not a satisfactory index of the *credibility* of the threat, and second because it is not a satisfactory index

of the *effectiveness* of the threat. Furthermore, the assumption that the threat is a significant deterrent factor is not required in order to provide at least two equally plausible alternative explanations of the data. This situation calls for further empirical research in order to choose among alternative interpretations of Russett's findings and in order to gain more knowledge about the determinants of national non-aggression.

Summary of Russett's Study

The sample consisted of 17 historical (1935-1961) deterrence situations, having in common the following elements:

(1) The "attacker" (A) and the "defender" (D) were both major powers—e.g., US, USSR, China, Germany (in several cases, D was a coalition of major powers).

(2) The "pawn" (P) was a minor power—e.g., Iran, Egypt, Cuba, West Berlin.

(3) At a particular time (t), A threatened to attack P militarily.

(4) At some other time, either before or after t but before A could carry out the attack, D overtly expressed an intention to defend P militarily (i.e., to counterattack A if A attacked P). This expressed intention is referred to below as "D's deterrent threat."

Russett stated the problem as follows:

. . . Propositions about factors which determine the credibility of a given threat need to be tested systematically on a comparative basis. . . . We shall explore the question of what makes a threat credible by asking which threats in the past have been believed and which disregarded [p. 98].

This implies that the 17 cases were to be subdivided into two groups according to whether or not A *believed* D's deterrent threat, and then compared in order to see

what factors are differentially associated with belief and disbelief. In fact, however, the cases were subdivided according to whether or not A subsequently *attacked* P. The implicit assumption involved here was that if A believes D's threat, A will not attack P, but that if A does not believe D's threat, A will attack P. On the basis of this assumption, the observed *effectiveness* of D's threat (success = A does not attack P, failure = A attacks P) was used as an index of its *credibility* (high credibility = A believes that D is likely to carry out his threat, low credibility = A believes that D is unlikely to carry out his threat). In six cases, the observed outcome was success, while in the remaining eleven cases, the observed outcome was failure.

Nine independent variables were observed, each of which was thought to be a determinant of credibility:

- (1) Relative importance of P, defined in terms of population size.
- (2) Relative importance of P, defined in terms of gross national product.
- (3) Existence, prior to the crisis, of a formal commitment by D to defend P.
- (4) Relationship between the overall strategic military capabilities of D and A.
- (5) Relationship between the local (in P's territory) military capabilities of D and A.
- (6) The type of political system possessed by D (dictatorship vs. democracy).
- (7) Military cooperation between D and P, in the form of arms and advisers sent to P.
- (8) Political interdependence between D and P.
- (9) Economic interdependence between D and P, defined in terms of the amount of trade between them relative to the average amount of trade each has with all other countries.

Although Russett did not specify which component of credibility would be affected by each of these variables, it appears that the first three and the last three are relevant to D's *willingness* to carry out the deterrent threat, while variables 4, 5, and 6 are relevant to D's *ability* to carry it out. It was assumed that A would take these factors into account in deciding whether or not to believe the threat. Each variable was dichotomized, and one of the two values was hypothesized to be associated with high credibility. The success and failure cases were then compared in order to test these nine hypotheses. The data for each case were presented in the appendix to Russett's paper (p. 109).

Russett concluded that none of the nine factors is a sufficient condition for high credibility, and that the specified values of variables 1, 2, 3, 4, 5, 6, and 8 are not necessary conditions for high credibility. The crucial conclusions were that the *necessary conditions for high credibility* are: (a) equality of military strength between D and A (on either variable 4 or 5), and (b) presence of military and economic interdependence between D and P (variables 7 and 9). This led to the following hypothesis: "If other factors are equal, an attacker will regard a military response by the defender as more probable the greater the number of military, political, and economic ties between pawn and defender" (p. 107). On the basis of these findings, a practical recommendation was made, to the effect that if the US wants to make more credible its promise to defend third areas, such as Europe, from Soviet attack, it can do so by increasing its economic dependence on those areas.

Credibility vs. Effectiveness

A crucial difficulty in Russett's argument arises from his use of the (observed) ef-

fectiveness of D's threat as a measure of its (unobserved) credibility. This assumes that in all cases of successful deterrence, D's threat was credible, and in all cases of unsuccessful deterrence, D's threat was not credible. Unless this assumption is correct, no valid inferences about the determinants of high credibility can be drawn from the data.

Clearly, the equation of effectiveness and credibility is not simply a matter of definition. Effectiveness is the *behavioral outcome* of D's threat, defined in terms of A's subsequent action toward P (attack or no attack). Credibility, however, refers to A's *cognitive reaction* (belief or disbelief) to D's threat. Since these two variables are not identical, there is no strictly logical justification for assuming a one-to-one correspondence between them.

One might rely on the empirical assumption that in order to be effective, a threat must be highly credible, i.e., that high credibility of D's threat is a *necessary* condition for successful deterrence. If this is true, then it is legitimate to infer from observed success that the threat was believed by A. However, one cannot also infer from observed failure that the threat was *not* believed by A. This can be maintained only if high credibility is also a *sufficient* condition for success. In other words, the general empirical assumption required as justification for using success and failure as an index of credibility is that: "A believes D's threat" is a necessary *and* sufficient condition for "A does not attack P."

But this assumption, in unqualified form, contradicts Russett's explicit theory for predicting A's response to the deterrent threat. Using the probability-utility model which is currently popular in the analysis of national decision-making (cf. Snyder, 1960;

Singer, 1962), Russett specified four variables that jointly determine A's behavior:

- (1) V_a = the utility to A of an attack on P which is *not* countered by D;
- (2) V_w = the utility to A of an attack on P which is countered by D;
- (3) V_o = the utility to A of not attacking P and consequently not being attacked by D;
- (4) s = A's subjective probability that an attack on P would not be countered by D.

The conditions under which D's deterrent threat will fail were then stated in the following formula:

Precisely, he (A) will press the attack only if:
 $V_a(s) + V_w(1-s) > V_o$ [Russett, p. 107].

$V_a(s)$ represents the *expected utility* (to A) of an attack on P which is not countered by D, while $V_w(1-s)$ represents the expected utility (to A) of an attack which is countered by D. (It is important to note that the term $1-s$ represents credibility, or A's subjective probability that D *would* counter an attack on P.) The algebraic sum of these two expected utility terms represents the *total expected utility of attacking* (EU_a). Thus, the theory states that A will attack P (and therefore, by definition, D's threat will fail) whenever $EU_a > V_o$. By implication, it also states that A will not attack P (i.e., D's threat will succeed) whenever $EU_a \leq V_o$. Thus, the statement that high credibility is a necessary and sufficient condition for success is equivalent to the statement that a high value of $1-s$ (e.g., between 0.5 and 1.0) is a necessary and sufficient condition for $EU_a \leq V_o$. That this is false can be shown by analysis of the formula.

First, note that the value of EU_a must always fall within the range defined by the values of V_a and V_w . This is so because EU_a is actually a weighted average of V_a and V_w , where the weights are provided

by the complementary probabilities, s and $1-s$. By formula, $EUa = Va(s) + Vw(1-s)$. If $s = 1$, then $1-s = 0$, and $EUa = Va$. As s decreases, $1-s$ increases, and the value of EUa shifts proportionately, reaching the arithmetic mean of Va and Vw when $s = 0.5 = 1-s$. As s decreases further, $1-s$ increases further, and the value of EUa moves closer and closer to the value of Vw , until $s = 0$, $1-s = 1$, and $EUa = Vw$.

Since EUa must fall between Va and Vw for all values of $1-s$, it follows that $EUa < Vo$ for any value of Vo which is greater than both Va and Vw , regardless of the value of $1-s$. Therefore, no value of $1-s$ is a necessary condition for $EUa \leq Vo$. Furthermore, $EUa > Vo$ for any value of Vo which is smaller than both Va and Vw , regardless of the value of $1-s$. Therefore, no value of $1-s$ is a sufficient condition for $EUa \leq Vo$. This implies that high credibility is neither necessary nor sufficient for success, according to Russett's own formula. Therefore, his predictive model does not justify the use of effectiveness as an index of credibility.

However, the formula does allow us to specify a set of conditions under which high credibility will be both necessary and sufficient for success. This will be the case whenever $Va > Vw$ and $Vo = \frac{1}{2}(Va + Vw)$. Under these conditions, EUa will be less than Vo if, and only if, credibility is high ($1-s > 0.5$). Since Vo falls exactly at the midpoint between Va and Vw , EUa can be less than Vo only if Vw is weighted more heavily than Va , and this can occur only if $s < 0.5$ and $1-s > 0.5$, since these are the respective weights of Va and Vw . It can easily be seen that if Vo is not equal to the arithmetic mean of Va and Vw , then the minimum value of $1-s$ that is required in order to make $EUa \leq Vo$ will be different from 0.5. If Vo is greater than the

mean, then the minimum required value of $1-s$ is proportionately reduced, since EUa will not have to come as close to Vw in order to be less than Vo . Conversely, if Vo is less than the mean, then the minimum required value of $1-s$ is proportionately increased, since EUa will have to come closer to Vw in order to be less than Vo . For example, consider all cases in which $Va = +1$ and $Vw = -1$. If $Vo = 0$, then any value of $1-s > 0.5$ will lead to $EUa < Vo$. But if $Vo = -0.2$, then any value of $1-s > 0.6$ will lead to $EUa < Vo$. And if $Vo = -0.5$, then any value of $1-s > 0.75$ will lead to $EUa < Vo$.

In general, for all values of Vo which lie between Va and Vw , where $Va > Vw$, the minimum magnitude of $1-s$ (degree of credibility) which is required in order to predict that A will not attack P (successful deterrence) is directly proportional to $Va - Vo$, or in other words, is inversely proportional to $Vo - Vw$. Thus, as Vo varies from Va to Vw , the minimum degree of credibility which will be associated with success varies from 0 to 1. If one does not wish to define high credibility as $1-s > 0.5$, it is possible to give a more general rule. High credibility will be both necessary and sufficient for successful deterrence whenever: (a) $Va > Vo > Vw$, and (b) high credibility is defined as any value of $1-s$ which is greater than $(Va - Vo) / (Va - Vw)$. Under these conditions, if the formula for predicting A's behavior is true, it is legitimate to infer from observed success that D's threat was credible, and to infer from observed failure that D's threat was not credible.

The above analysis reveals the precise points at which Russett's study needs to be strengthened. First, the meaning of high credibility and low credibility needs to be more clearly defined. This is essential if

the inference from outcome to credibility is to mean the same thing in each of the 17 empirical cases. But let us make the reasonable assumption that the less exact terms "likely" and "unlikely" imply a cutting point somewhere near $1-s=0.5$. Now, since credibility is only one determinant of behavioral outcome, one cannot infer the degree of credibility directly from outcome except under specified boundary conditions. In order for success to be a valid indicator of high credibility, and failure a valid indicator of low credibility, V_a must be greater than V_w , and $V_a - V_o$ must equal approximately $\frac{1}{2}(V_a - V_w)$, for all observed cases. It is not necessary that the absolute values of V_a , V_o , and V_w be identical across all 17 cases, but only that their relative values be the same for all cases. The critical weakness lies in the fact that Russett presented no evidence or argument to support the assumption that, in each case, $V_a - V_o = \frac{1}{2}(V_a - V_w)$. If we wish to accept this assumption, then we can accept conclusions about credibility based on observations of the effectiveness of D's threat. But if there is no reason to accept this assumption, then there is no reason to accept any conclusions about the determinants of credibility.

Apparent vs. Actual Effectiveness

A second major difficulty arises from the fact that effectiveness itself was not actually observed. In summarizing Russett's study, I stated that he defined success as the behavioral outcome "A does not attack P," and failure as the outcome "A attacks P." Strictly speaking, this was an oversimplification. Russett's explicit definition of success was "an instance when an attack on the pawn is prevented or repulsed without conflict between the attacking forces and regular combat units of the major power 'defender'" (p. 98). The crucial

element in this definition is that D's threat *prevents* A from attacking P. But prevention implies more than the simple observation that A does not attack; it also implies that D's threat is crucial in producing that outcome, i.e., that the threat was an *effective inhibitor of an action that otherwise would have occurred*. According to this meaning of effectiveness, D's threat can be considered successful only if: (a) in the absence of the threat, A would have attacked P; (b) nothing else besides D's threat acted as an effective deterrent; and (c) A did not attack P. If A would not have attacked anyway, then the same behavioral outcome only gives the *illusion* that D's threat was effective; this case can be called *apparent* effectiveness. Unless both conditions *a* and *b* can be assumed to hold, then conclusions about the determinants of actual effectiveness cannot confidently be derived from the data.

The crucial assumption is that in the absence of D's threat (or some other deterrent factor), A would have attacked P. This was assumed to be true for each of the six "success" cases. Apparently, the only basis for this assumption is the fact that in each case, A had threatened to attack P. But to infer from A's threat that A will attack P unless prevented by some deterrent is to attribute perfect credibility to A's threat. One cannot ignore the possibility that A might threaten P, not because A actually intends to attack P, but rather as a means to some other end. For example, it might be simply a tactical move on the part of A in order to distract D's attention, or in order to win a concession on some other point. Possibly the only factor which is needed in order to differentiate the "success" and failure cases is that in the former, A did not intend to attack P, while in the latter, A did intend to attack.

Even if it is granted that A would have attacked, it still must be shown that it was D's deterrent threat and not some other input which made the difference. Suppose, for example, that D's deterrent threat was only one of several world social forces that were brought to bear on A in response to A's threat to attack P. Various countries, international organizations, or even factions within A may have tried, by various means (including threats of nonmilitary retaliation, promises, and suasion), to deter A from such aggression. The total pattern of such forces might have to be considered the effective deterrent in this case, and it is conceivable that D's threat alone would not have been sufficient. The implication of the above argument is that before one can accept Russett's data as evidence regarding the determinants of threat effectiveness, evidence must be obtained on the six "success" cases to show that they are in fact cases of actual success.

Alternative Interpretations of the Data

If the preceding arguments are correct, it follows that Russett's data can, by themselves, tell us nothing about the determinants of either the credibility or the effectiveness of D's deterrent threat. However, the data are not meaningless, since the observed dependent variable is itself interesting. There were, in fact, two different behavioral outcomes (attack vs. no attack) within the set of 17 deterrence situations studied. If one admits the possibility that D's threat had little or no causal relevance to the outcome, then the data can be re-examined in an effort to develop hypotheses about the determinants of A's final response.

To facilitate the analysis, the findings can be summarized by means of a series of 2×2 contingency tables. This can be done readily from Russett's appendix (p.

109), since all variables there were dichotomized. Table 1 shows the number of "attack" (failure) cases and the number of "no attack" (success) cases for each value of the nine independent variables. To provide a rough measure of the degree of association between the independent variables and behavioral outcome, phi coefficients have been computed, and the variables are listed in descending order according to their corresponding phi values.

One may have some reservations about applying a statistical test to these data, on grounds that historical connections among many of the cases contradict the assumptions of independence required by the test. For example, three of the "success" cases involved the US in the role of defender against the USSR, all within a two-year period; and four of the failure cases involved Great Britain and France in the role of defender against Germany, again within a two-year period. It can be argued that these seven cases are not really seven independent cases, but rather only two, and that the sample is therefore artificially inflated from the point of view of statistical inference. However, the effect of reducing sample size would be to reduce the degree of confidence one could have that any of the correlations are different from zero. Recognizing that the purpose of the present analysis is not to test hypotheses, but rather to suggest them, it does not seem necessary to apply the most rigorous statistical criteria. Thus, for each 2×2 table, the chance probability of obtaining the given array of data (or a more extreme array) has been computed by means of Fisher's exact test, simply as a basis for comparing the obtained correlations.

Among the nine independent variables, economic interdependence between D and P is most highly associated with outcome.

TABLE 1

BEHAVIORAL OUTCOME OF 17 DETERRENCE SITUATIONS AS A FUNCTION OF NINE SITUATIONAL VARIABLES

Situational variable		Number of cases for each outcome		Phi coefficient	Probability by Fisher's Exact Test
		Attack	No attack		
(9) P-D Econ. Interdependence:	Present	2	5	0.63	0.076
	Absent	9	1		
(7) P-D Military Cooperation:	Present	7	6	0.41	0.140
	Absent	4	0		
(8) P-D Polit. Interdependence:	Present	4	4	0.28	>0.200
	Absent	7	2		
(4) Strategic Superiority, D/A:	Present	5	4	0.20	>0.200
	Absent	6	2		
(6) D's Political System:	Dictatorship	2	2	0.17	>0.200
	Democracy	9	4		
(5) Local Superiority, D/A:	Present	2	0	-0.27	>0.200
	Absent	9	6		
(3) Precrisis Commitment by D:	Present	6	1	-0.37	0.160
	Absent	5	5		
(1) Population Ratio, P/D:	20%+	5	0	-0.48	0.075
	20%-	6	6		
(2) GNP Ratio, P/D:	5%+	5	0	-0.48	0.075
	5%-	6	6		

On the basis of this finding, Russett argued that the credibility of D's threat is positively associated with D-P interdependence. But the obtained correlation can be interpreted in quite a different way. Perhaps in most cases A threatened to attack a pawn which was known to be important to D *in order to harass D* and perhaps win some other point in another arena of A-D interaction. If so, then A's threat and D's deterrent threat set up an apparent deterrence situation, which was bound to result in no attack because A had no intention of doing so. Thus, D-P economic interdependence may be associated with no attack, not because it influenced A's decision about attacking *following* D's threat, but rather because it influenced A's decision to make a false threat against P *prior* to D's deterrent threat.

Although military cooperation between

D and P was more weakly associated with outcome, it was nevertheless present in all cases of no attack. This may be accounted for in the same way as the previous correlation. The fact that D sent arms and advisers to P may have been treated by A as evidence that D was interested enough in P to be affected by a threat against P. This objective fact of D-P cooperation may have been the trigger for a false threat by A rather than a cue enhancing the credibility and effectiveness of D's deterrent threat. This same line of argument is consistent with the fact that the relative military strength of A and D was not associated with outcome: if A did not really intend to attack P, then it would not be necessary for D to be militarily superior to A in order for the outcome to be "no attack."

It is more difficult to account for the

TABLE 2
 BEHAVIORAL OUTCOME OF 17 DETERRENCE
 SITUATIONS AS A FUNCTION OF DATE OF
 OCCURRENCE OF THE CRISIS

Date of crisis	Number of cases for each outcome		Phi coefficient	Probability by Fisher's Exact Test
	Attack	No attack		
Post-1945	4	6	0.62	0.017
Pre-1940	7	0		

unexpected finding that the relative size of P and D was negatively correlated with outcome. Neither the "credibility" interpretation nor the "false threat by A" interpretation is consistent with this result, since a P that is relatively unimportant should neither be selected as a special target by A nor produce a highly credible threat by D. Perhaps this merely reflects the fact that the size of P relative to the defender does not function as a sign of the importance of P in D's value hierarchy. But if this is true, why shouldn't the correlation be near zero rather than near -0.50 ? One answer might be that the sample is so small that one cannot attribute much significance to the obtained correlation. However, this would remove the necessity for interpreting any of the obtained correlations, since the statistical significance level obtained by Fisher's exact test is 0.075 for both the most positive and the most negative correlations in the table. A second possible answer is that the P/D ratio varies mainly because of variations in the size of D rather than because of variation in the size of P. A glance at Russett's list of cases (1963, p. 99) shows that in all six "success" (no attack) cases, D was either the US or the USSR, which are much larger than other major powers, while in the failure (attack) cases, more than half involved Great Britain and France as defenders.

When D is very large, P is more likely to be relatively small. This difference alone can account for the obtained negative correlation between relative size of P and behavioral outcome.

Differences in the size of D are coupled to another fact which may be of significance. One can also see from Russett's list of cases that all of the "no attack" outcomes occurred during the post-World War II era, while most of the "attack" outcomes occurred prior to the war. Table 2 shows the relationship between behavioral outcome and date of occurrence of the crisis.

The striking fact about these data is that the historical period is as strongly associated with outcome as is D-P economic interdependence. But the correlation between date and outcome permits an entirely different interpretation. It is conceivable that a crucial change has taken place in the total international situation, sharply increasing the total pattern of deterrent pressures which can be brought to bear on a potential attacker. Perhaps the focusing of world political pressures through the United Nations exerts a great restraining force, a general downward pressure on *EUa*, even without explicit deterrent threats by D. It is quite plausible that such extra pressures were far weaker during the prewar period, so that the total deterrent environment has changed. Under this interpretation, the association between outcome and D-P economic or military interdependence may be merely a reflection of the fact that the postwar policies of both the US and the USSR have led to an increase in the number of satellites for each. In this case, P-D interdependence can be seen as a cause for D's deterrent threat, but as causally irrelevant to the behavioral outcome.

A different mode of analysis of the

data, which Russett apparently used in arriving at several of his conclusions, is to look for factors which may be necessary conditions for the occurrence of the preferred outcome (no attack). Since a necessary condition is one which must be present in order for the desired event to take place, it follows that in order to claim that a factor may be necessary, it must be present in all cases with that outcome. From Tables 1 and 2, it can be seen that the following conditions were present in all six cases where the outcome was "no attack": (1) there was military cooperation between D and P; (2) D did not possess local military superiority over A; (3) the population ratio of P/D was less than 20 percent; (4) the GNP ratio of P/D was less than 5 percent; (5) the crisis occurred after 1945. It is also true, although not shown in these tables, that in each case: (6) D was either the US or the USSR; (7) D possessed nuclear weapons at the time of the crisis.

Thus it is possible that any one of these conditions, or any combination of two or more of them, is a necessary condition for the outcome to be "no attack." There are probably many other factors which were also present in each of these cases, so that a decision as to which, if any, are necessary to the outcome is not possible on the basis of the limited data which have been presented.

Implications for Further Research

The alternative interpretations which I have given for Russett's data reflect a theoretical approach to the study of deterrence which is generally similar to his. The differences seem to lie mainly in which factual assumptions we are willing to make or which potentially relevant variables we are willing to ignore for purposes of a particular study. Thus, the general theory

which we both use postulates that if there are factors motivating A to attack P, then some inhibiting factors must be present in order to prevent that attack. Russett's study concentrated on a specific subset of all possible inhibiting factors, while assuming that other inhibiting factors as well as the motivating factors were constant or else could be ignored. My thesis is simply that greater attention must be paid to the latter two sets of factors if valid conclusions are to be drawn about the first set.

The model I have been using predicts behavioral outcome from the relative strength of two motives: A's motive to attack P (Ma) and A's motive not to attack P (Mo). It states that A will attack if $Ma - Mo > 0$, and that A will not attack if $Ma - Mo \leq 0$. Further, Mo is broken down into at least two components: the motivation not to attack which arises from D's deterrent threat (Dd), and the motivation not to attack which arises from other deterrent factors (Od). $Mo = Dd + Od$. Thus, A will attack if $Ma - Dd - Od > 0$, but will not attack if $Ma - Dd - Od \leq 0$. The magnitude of Dd will make a difference to the outcome only if $Ma - Od > 0$, because the predicted outcome will always be "no attack" when $Ma - Od \leq 0$.

The "other inhibiting factors" which contribute to the strength of Od need not be restricted to threats of military retaliation by nations other than D. It is possible to broaden the concept of deterrence to include other methods of influencing A, such as threats of economic retaliation or promises (by D or by other nations) of reward for doing something else (cf. Milburn, 1959, p. 139). In addition, inhibiting factors need not all arise in A's social environment, but may come partly from politically active groups within A. Thus, one may consider both negative and posi-

tive pressures against attacking, arising from various sources.

Future empirical research on the determinants of effective deterrence must deal explicitly with these various classes of variables, since it cannot be assumed *a priori* that they are negligible in any historical case. If the problem is to determine whether some particular variable influences the magnitude of Dd , and if the behavioral outcome of the deterrence situation is to be used as an index of the magnitude of Dd , then ideally the cases would be selected so that they all have the same magnitude, greater than zero, of $Ma - Od$. If it is not possible to find a sufficient number of such cases on record, then it may be possible to do the next best thing, which is to measure the strengths of the various component motivational forces and make empirical predictions by plugging the obtained values into one or more predictive formulas in order to see whether predictions are improved by including the variable in question. If the problem of measuring the motivational forces appears insurmountable for the present, it may be necessary to give up the motivational model as a basis for designing the study, or even to redefine the problem.

One approach which does not depend so directly on the motivational model starts with the postulate that the behavioral outcome of the deterrence situation is determined by the total pattern of communicative influence attempts (threats, promises, and suasion) directed at A's decision-makers after A threatened to attack P. It can be assumed that some of these messages will be favorable to the attack, and that others will oppose it; it can also be assumed that some countries will not attempt to influence A in a particular case, thus by default communicating permissiveness. It can also be assumed that the impact of each message

will be a function of the power of its sender; thus each message can be weighted according to some index of the sender's power, and a weighted sum of favorable, unfavorable, and permissive messages can be obtained. A's response can be predicted from the ratio of the total strength of unfavorable messages to the total strength of favorable messages, perhaps modified by the total amount of permissiveness present in the situation. The link between these factors and the motivational model is contained in the assumption that Mo is positively correlated with the total strength of unfavorable messages, and that Ma is increased by the occurrence of favorable messages and perhaps by the presence of permissiveness. Whether it is easier to gather data on the matrix of influence attempts impinging on A, or directly on the motivational state of A, will have to be answered in the process of trying to gather both kinds of data.

An important implication of the present discussion is this: the question of whether deterrent threats of military retaliation are *ever* effective in preventing aggression is prior to the question of what determines the effectiveness of such threats. Can one find evidence to refute the null hypothesis that the incidence of military attacks by one nation on another, following a threat by the attacker, is the same whether or not a deterrent threat is made? Perhaps this hypothesis can be tested by broadening the empirical base beyond deterrence situations involving only the defense of small third parties by large defenders against large attackers. By including self-defense situations as well, and by taking all cases regardless of the size of the principal participant nations, a much larger sample of deterrence situations can be obtained. With a larger sample, the chances will be greater of finding cases in which there was no deterrent threat

(remember that there apparently was such a threat in every case in Russett's sample); unless such cases can be found, it will not be possible to test the hypothesis in question.

Russett's study was an effort to apply *ex post facto* experimental logic to the empirical investigation of deterrence. I have tried to show some of the difficulties encountered in such studies. But despite these difficulties, I believe that this kind of systematic analysis of historical cases can lead to an advance in knowledge about international relations. If nothing more, it can at least help clarify the nature of and the reasons for our ignorance of general laws governing the behavior of nations.

REFERENCES

- KAUFMANN, WILLIAM W. *The Requirements of Deterrence*. (Memo No. 7.) Princeton, N.J.: Center for International Studies, 1954.
- MILBURN, THOMAS W. "What Constitutes Effective Deterrence?" *Journal of Conflict Resolution*, 3, 2 (June 1959), 138-45.
- RUSSETT, BRUCE M. "The Calculus of Deterrence," *Journal of Conflict Resolution*, 7, 2 (June 1963), 97-109.
- SCHELLING, T. C. *The Strategy of Conflict*. Cambridge: Harvard University Press, 1960.
- SINGER, J. DAVID. *Deterrence, Arms Control and Disarmament: Toward a Synthesis in National Security Policy*. Columbus, Ohio: Ohio State University Press, 1962.
- SNYDER, GLENN H. "Deterrence and Power," *Journal of Conflict Resolution*, 4, 2 (June 1960), 163-78.