Review Article: When Do Deterrent Threats Work?
JACK S. LEVY

This review is concerned with the general theoretical question of when deterrence works. Under what conditions are deterrent threats successful in persuading a potentially hostile adversary not to take certain actions that would be contrary to one's own interests, and under what conditions do deterrent threats fail to modify the adversary's hostile intentions or even provoke him into taking the undesired actions? I begin by identifying three distinct bodies of theoretical literature on deterrence. I then focus more specifically on the quantitative empirical literature on deterrence and the answers it provides to the question of when deterrence works.

ALTERNATIVE PERSPECTIVES ON DETERRENCE

Deterrence can be defined generally as 'the persuasion of one's opponent that the costs and/or risks of a given course of action he might take outweigh its benefits.' Statesmen have long sought to deter their adversaries from taking hostile actions, but it is only in the nuclear era that scholars have attempted to theorize systematically about deterrence, and to identify the conditions under which deterrence threats will succeed in dissuading an adversary from taking actions that are harmful to one's own interests.

There are three distinct bodies of literature which address the general theoretical question of when deterrence works. The first might be described as the 'classical theory of deterrence.' Developed in the fifteen years after the Second World

Department of Political Science, University of Minnesota. This research has been supported by the Stanford Center for International Security and Arms Control, the Carnegie Corporation and by a Social Science Research Council/MacArthur Foundation Fellowship in International Peace and Security. The views expressed here do not necessarily represent those of the Center, Council or either Foundation. An earlier version of this study was presented at a November 1986 Workshop on Deterrence sponsored by the US National Academy of Sciences/National Research Council Committee on Contributions of Behavioral and Social Science to the Prevention of Nuclear War. A revised version of the original study expanding on these themes will appear in Perspectives on Deterrence, edited by Paul Stern, Robert Axelrod, Robert Jervis and Roy Radner (New York: Oxford University Press, 1989). This research has benefited from the helpful comments and criticism of Christopher Achen, Robert Jervis, Kimberly Marten, John Mearsheimer, Paul Stern, Steven Weber and especially Alexander George.

2 Theorists as far back as Thucydides dealt with this question, though they did not use the specific term and did not attempt to integrate their ideas into a systematic body of knowledge.
War, this is a rational theory of deterrence which provides powerful and elegant explanations for many dimensions of interstate behaviour based on an axiomatic system of deductive logic. Classical deterrence theory assumes that states can be treated as unitary actors who have a well-defined set of interests, who act to maximize those interests given existing constraints, and who do this through rational decision-making processes based on conscious cost–benefit calculations.3

Although there are several different variations of rational deterrence theory, there is a consensus regarding some minimum conditions required for deterrence to work. A state must clearly define its commitment to defend a particular interest, communicate that commitment to the potential aggressor, possess a sufficiently potent military capability to impose costs on the adversary which exceed his expected gains, and demonstrate its resolve to implement the threat even though there may be short-term costs to itself. These theoretical requirements for deterrence have had a profound impact upon US strategic doctrine and security policy in the last thirty years.4

In the last decade or so there has arisen a second body of literature which poses an important challenge to the classical theory of deterrence. These critics are distinguished from classical theorists on both theoretical and methodological grounds. They argue that statesmen are often driven to challenge the deterrence commitments of their adversaries in spite of the clarity of the commitments, the credibility of the threat and the military capabilities available to support the threat. They reject the unitary actor assumption, and argue that political leaders are often driven by bureaucratic or domestic political interests to use military force in spite of existing deterrent threats. They also reject the rationality assumption, and draw on the literature in organizational theory and social psychology to explain why statesmen do not always engage in the type of information search and careful cost–benefit calculations required by rational expected-utility theory. Moreover, whereas the classical theorists proceed deductively and make little effort to subject their propositions to empirical test, these other scholars generally proceed inductively and rely heavily on intensive case studies of state behaviour in deterrence situations. They conclude that states frequently do not act consistently with the predictions of classical deter-

3 There is some debate over whether formal deterrence theory requires that states actually make rational cost–benefit calculations or whether they only act as if they behaved rationally. See Bruce Bueno de Mesquita, The War Trap (New Haven, Conn.: Yale University Press, 1981), pp. 29–33; Christopher H. Achen and Duncan Snidal, 'Rational Deterrence Theory and Comparative Case Studies', World Politics, forthcoming.

rence theory. Much of the case-study literature on deterrence is also very concerned with the policy implications of theory, and argues that a military doctrine based on the flawed assumptions of rational deterrence theory runs unacceptable risks in the nuclear age.5

There is still another body of literature which addresses the question of when deterrence works. This is distinguished by its methodological approach rather than theoretical orientation, and belongs to the tradition of quantitative empirical research in international relations. Its aim is to construct hypotheses regarding the conditions for deterrence success and failure, to operationalize those hypotheses with quantitative indicators of the key theoretical concepts, and to subject those hypotheses to systematic empirical test based on large-N aggregate studies. Most of this work has employed aggregate data methods, but events data have also been utilized.6

Most of the hypotheses investigated by the quantitative empirical literature are consistent with the Realpolitik assumptions underlying classical deterrence theory. The key variables include the dyadic balance of military power, alliance patterns, political commitments, trade patterns, and other variables reflecting state interests, capabilities and commitments. The quantitative literature is more concerned with the hypotheses themselves and with the empirical validity of those hypotheses, than with the formal theoretical apparatus from which they are derived. There is no reason in principle why the psychological or domestic political concerns of the case study literature cannot be incorporated into empirical models of deterrence, and in fact the most recent quantitative empirical studies have done precisely that.7

Both classical deterrence theory and the case study literature on deterrence will be familiar to most readers, and there is little need for a comprehensive survey of that literature here. Quantitative empirical studies of deterrence are much more recent and less familiar, however, and have not been reviewed in any systematic way. Consequently, this article will focus primarily on the quantitative empirical literature on deterrence, with particular emphasis on the conditions


7 Although these hypotheses are ‘consistent’ with the rational and unitary actor assumptions of classical deterrence theory, they have not been formally derived from that theory, and in fact are more specific than those suggested by classical deterrence theory. Although classical deterrence theory provides a framework for the analysis of interests, power and costs, it rarely specifies precisely what those interests are. The quantitative empirical literature on deterrence goes beyond classical theory by attempting to evaluate the relative importance of various Realpolitik variables, along with some others, in determining when deterrence works.
under which deterrent threats are likely to succeed. In examining this literature I will attempt to place its major assumptions and findings in the context of the broader theoretical and empirical literature on deterrence. I will note where the findings of the quantitative studies reinforce those from the case study literature and where those findings are inconsistent.

In spite of its importance, this question of when deterrence works has not received extensive treatment in the quantitative literature on international relations. The research most directly related to this question is that of Huth and Russett on extended deterrence, following up on Russett’s influential 1963 study, and their work will be examined in detail. There has also been some interesting research on military threats and crisis behaviour which does not focus specifically on deterrence but which has important implications for the question of the effectiveness of deterrent threats. This includes some of the work by North and his colleagues on the 1914 project; by Singer, Maoz, Leng and others on recent extensions of the Correlates of War project; and by Karsten et al. in their historical study of military threats. I will examine these studies here and attempt to draw out their implications for the effectiveness of deterrence threats.

The empirical findings from these studies cannot be easily summarized. They are sensitive to the empirical domain of the analysis, the selection of cases, the operational indicators of the independent and dependent variables, and to other aspects of the research design. Consequently, a meaningful analysis requires the specification of the theoretical questions towards which each study is directed and a discussion of the methodology by which it is carried out. For this reason I will examine a few important studies in some detail rather than present a laundry list of the empirical results of every study having something to do with deterrence.

One important theme in the literature concerns the relative importance of the dyadic balance of military capabilities in deterrence crises, and I will begin with that question. I will then examine in greater detail the ongoing research program of Russett and Huth, which deals with a wide range of hypotheses regarding extended deterrence. Other empirical literature will be discussed where it has an important bearing on these studies.

Before we begin it is necessary to make an important theoretical distinction between ‘general deterrence’ and ‘immediate deterrence’. Immediate deterrence refers to the relationship between opposing states ‘where at least one side is seriously considering an attack while the other is mounting a threat of retaliation in order to prevent it’, whereas general deterrence refers to adversaries who ‘maintain armed forces to regulate their relationship even though neither is anywhere

---


9 Specific studies will be cited as they are examined. Because many of these studies were not designed specifically and solely to answer the theoretical question of the conditions affecting the success or failure of deterrence, my analysis of the bearing of these studies on that question should not be interpreted as a judgement of their overall merit.
near mounting an attack'. Immediate deterrence assumes the existence of a crisis or dispute in which military threats have been made and hence assumes the existence of the underlying conditions generating the crisis. General deterrence makes no assumption that there exists a crisis or serious dispute, that one state is seriously considering an attack on the other, or that a specific deterrent threat has been issued. Whereas general deterrence is concerned with the sources of crises as well as the conditions for crisis stability (stability within a crisis), immediate deterrence is concerned only with crisis stability. This study, and in fact most of the empirical literature on deterrence, is concerned with immediate deterrence rather than general deterrence, because the very question of the conditions under which deterrent threats are effective assumes the prior existence of a threat. As we will see, however, this distinction between general and immediate deterrence raises a difficult analytical problem.

THE ROLE OF CAPABILITIES

In spite of the numerous preconditions for successful deterrence, there are some proponents of a 'power politics' model who focus primarily on military capabilities alone as the central element of deterrence. They assume that a state will not initiate a war that it expects to lose, so that the defender's possession of superior military capabilities (along with the adversary's recognition of that superiority) is a sufficient condition for deterrence. This proposition is reflected in the old adage *si vis pacem para bellum* (if you want peace prepare for war), which for centuries has been used by statesmen to justify the expansion of their armaments programmes.

This proposition seems to imply sufficient rather than necessary conditions for deterrence: it does not say that the strong will always attack the weak, but only that the weak will never attack the strong (unless they have help from allies). Some adopt a stronger version of the hypothesis, however, and suggest that superior capabilities are a necessary as well as sufficient condition for deterrence, that the strong will attack the weak if there is nothing to prevent them from doing so. This is implied by the Athenians' argument to the Melians that 'the strong do what they have the power to do and the weak accept what they have to accept'.

There are enough situations in which the strong do not attack the weak in spite of their ability to do so, as well as logical flaws in the argument, to cast serious doubt on the stronger version of the peace-through-strength hypothesis. Even the weaker version of the hypothesis, however, is beset by logical problems in addition to some important contradictory cases. One problem is that it fails to incorporate the interests of the actors involved in the conflict. It is reasonable

---

to assume that states consider the likely costs and benefits from war as well as the probability of victory based on the balance of dyadic capabilities, so that actions involving a less-than-even odds of winning can be rationally undertaken if their potential benefits are sufficiently great and if there are some limits on the costs of defeat.\(^{12}\) In addition, the costs and benefits of war should be compared to those of other alternatives, including the alternative of doing nothing. Consequently, weaker states may rationally initiate war if they believe that the existing status quo is so unacceptable that they have nothing to lose, or that an attack by the adversary is imminent and that there are advantages in striking first.

In spite of these theoretical problems, the power politics hypothesis of deterrence has numerous advocates. The popularity of the hypothesis is suggested, perhaps, by the number of empirical studies designed to test it.\(^{13}\)

The North 1914 Studies

A fairly early study that has some bearing on the question of the importance of capabilities for the effectiveness of deterrent threats is the 1961 article by Zinnes, North and Koch, ‘Capability, Threat, and the Outbreak of War’.\(^{14}\) Part of North’s 1914 Project, this study uses content analysis of a fairly complete set of official diplomatic documents to examine the hypothesis that a state will not initiate a war if it perceives its (or its coalition’s) military capabilities to be ‘significantly’ inferior to those of its adversary.

The authors argue, with support from well-selected statements from high-ranking German and Austro-Hungarian political and military decision makers, that ‘both Austria and Germany possessed evidence of their own inadequate

\(^{12}\) The importance of the balance of interests at stake in a conflict is emphasized by George and Smoke in their concept of asymmetry of motivation and by Jervis in his concept of ‘intrinsic interests’. Both the probability of victory or defeat and the value of those outcomes have been integrated into an expected utility theory of war by Bueno de Mesquita. See George and Smoke, *Deterrence in American Foreign Policy*; Robert Jervis, ‘Deterrence Theory Revisited’, *World Politics*, 31 (1979), 314–17; Bueno de Mesquita, *The War Trap*.


capabilities – yet they were not deterred’. Frequency counts of key indicators reveal that perceptions of (and, the authors imply, concern with) the adversary’s hostility far exceeded perceptions of relative capabilities. The authors therefore reject the hypothesis that a state’s military inferiority is sufficient to deter its initiation of a war, and argue instead that if perceptions of threat or hostility or general dissatisfaction are ‘sufficiently’ great they can offset perceptions of insufficient capabilities.

In spite of some methodological problems relating to the validity of indicators based on a content analysis of official diplomatic documents, this is an important study. It was one of the first systematic empirical analyses to contradict the common assumption that perceptions of one’s inferior capabilities will preclude one from initiating a war, that superior capabilities are always sufficient to deter an adversary from attacking. Its findings gain additional support from other studies which employ different methodologies to investigate similar theoretical questions.

The ‘Correlates of War’ Studies

Some of the work by Singer and his colleagues on the Militarized Interstate Dispute (MID) project (an extension of Singer’s Correlates of War project) is relevant to the question of whether military superiority is sufficient for deterrence in international disputes, although they do not focus directly on deterrence per se. Maoz attempts to test a ‘capability model’ versus a ‘resolve model’ for all interstate disputes since the Congress of Vienna, using the dispute data he collected with Gochman. The capability/threat model predicts that the probability of victory for an initiator of a dispute (defined as ‘that state which has first committed a military confrontation action against another state’) varies positively with

---

16 Zinnes, North and Koch, ‘Capability, Threat, and the Outbreak of War’, p. 4. Note that unless the qualifier ‘sufficiently’ is operationally defined, its inclusion seriously reduces the explanatory and predictive power of the hypothesis by making it nearly non-falsifiable. These findings gain further support from North’s subsequent study, which uses events’ data on military actions (n = 354) as well as perceptual data from the documents. Robert C. North, ‘Perception and Action in the 1914 Crisis’, Journal of International Affairs, 21 (1967), 103–22.
the ratio of the initiator's capabilities to the target's capabilities.\(^{19}\) Capabilities are measured using the military expenditure and military personnel indicators from Singer's Correlates of War capability data.\(^{20}\) The resolve model predicts that the outcome of the dispute is determined primarily by the resolve of the two parties.\(^{21}\)

Maoz confirms the findings of Bueno de Mesquita and others that initiators are disproportionately likely to emerge victorious in militarized disputes. His central empirical result is that the initiator's success is unrelated to the balance of military capabilities between initiator and target (or between the initiator's coalition and the target's coalition), thus disconfirming the predictions of the capability/threat model. The various resolve indicators, on the other hand, are significantly associated with dispute outcomes: initiators tend to win when they display higher levels of hostility than their adversary or when they maintain control over the escalatory sequence of the dispute, lose when targets are more hostile or maintain the initiative, and tie when the two are matched on these indicators of resolve. Moreover, the outcomes of those disputes which escalate to war is determined more by the resolve indicators than by the capability indicators. Maoz concludes that 'initiators of serious interstate disputes tend to emerge as victors not because they are stronger than targets but because they are able to demonstrate that the stakes of the dispute are more important to them than to their opponents'.\(^{22}\)

It should be emphasized that Maoz's findings suggest that the motivations and actions of the initiator of the dispute are more important than those of the defender in determining the outcome. This runs against the standard focus in classical deterrence theory on the importance of commitment, resolve and signalling on the part of the defender. It is more consistent with the emphasis of George and Smoke on initiation theory and with a similar emphasis in the work of Lebow and of Jervis, Lebow and Stein.\(^{23}\)

\(^{19}\) I use 'capability/threat model' to refer to Maoz's 'threat model'. Maoz suggests another capability model, a power transition model, for which the key independent variable is change in relative military capabilities. This is a key variable in the process leading to war, but is not normally included in the theoretical or empirical literature on deterrence and will not be examined here. See Jack S. Levy, 'Declining Power and the Preventive Motivation for War', World Politics, 40 (1987), 82–107. Note that whereas the North project focused on perceptions of capabilities, Maoz focuses on 'objective' capabilities.


\(^{21}\) There may be some questions regarding Maoz's resolve measure, which is based on the relative levels of hostility reached during the dispute and on the extent to which one side undertakes 'incidents' to maintain the initiative during the dispute.

\(^{22}\) Maoz, 'Resolve', p. 221; Bueno de Mesquita, The War Trap. An important exception is that capability ratios do have a significant impact on the outcome of disputes and wars between major powers, though it is not indicated in the article whether the resolve indicators have an even greater impact. Maoz cautions, however, that whether this imbalance of resolve is real or whether it derives from the deceptive manipulation of risks by initiators needs to be analysed.

\(^{23}\) George and Smoke, Deterrence in American Foreign Policy; Lebow, Between Peace and War; Jervis, Lebow and Stein, Psychology and Deterrence. Blechman and Kaplan also find, in their study
Although Maoz provides a careful study of the role of capabilities and resolve in international disputes, it is necessary to point out some limitations on the study's relevance for the question of immediate deterrence (which, to repeat, was not its primary purpose), for these same considerations limit the applicability of numerous studies of capabilities, threats and resolve to the question of immediate deterrence. First, although all of the MID cases do involve situations in which at least one side has made a military threat against the other, they do not necessarily involve any specific attempt by the defender to deter the initiator from taking a specific action, so we could not necessarily conclude that any success by an initiator is necessarily a failure of deterrence. Second, Maoz does not differentiate between deterrence and compellence. Compellence refers to the use of threats to induce an adversary to do something or to stop doing something rather than to refrain from doing something he has not yet done, and is thus harder to implement than deterrence.\(^2^4\) Consequently, the failure of compellence does not necessarily imply the failure of deterrence. Third, Maoz's operational indicators of resolve are highly questionable and fail to tap the importance of the stakes involved or the intensity of actors' commitments. Finally, Maoz's definition of success and failure is not appropriate for the analysis of the success or failure of deterrent threats. A dispute which escalates to war is coded as a success for that side which wins the war militarily.\(^2^5\) This may be useful for the theoretical questions he is asking, but from the perspective of deterrence such an outcome should be treated as a failure.

Wayman, Singer and Goertz look at the impact of capabilities on the outcome of disputes from a slightly different perspective, and focus only on disputes between the major powers (\(n = 101\)). Their independent variables include both overall capability ratios, defined in terms of the COW project's six indicators of the demographic, economic/industrial and military dimensions of national strength, and also the allocation of resources between the industrial and military sectors. Although the direct relevance of their findings for deterrence is limited by the same factors mentioned above with respect to Maoz's study, the findings are none the less interesting. Industrial capabilities are more important than military preparedness (operationalized in terms of military expenditures and number of armed forces personnel) in determining the outcome of wars. Moreover, initiators who over-allocate in terms of expenditures \textit{vis-à-vis} the industrial base have been defeated in war more often than they have been victorious. The value of military superiority in war (as long as it does not come at the expense of industrial might) disappears for disputes that do not escalate to war: the weaker

\(^{24}\) Thomas C. Schelling, \textit{Arms and Influence} (New Haven, Conn.: Yale University Press, 1966), Chap. 2; David Baldwin, 'Power Analysis and World Politics', \textit{World Politics}, 31 (1979), 162–94.

party in terms of both personnel and expenditures tends to be more successful. The authors conclude by questioning the accuracy (at least for the post-Vienna period) of the statement by Frederick the Great that 'God is always with the strongest battalions'.

In another study which uses the MID data but which is not part of the Correlates of War project, Bueno de Mesquita compares the relative predictive power of a power politics model with that of his expected utility model, which attempts to incorporate the interests of states (as measured by their alliance commitments) as well as their military capabilities. He examines the proportion of war initiators having superior military capabilities and the proportion having positive expected utility for the 1816 to 1974 period. Although the predictions of these models often overlap, under certain conditions they diverge and consequently the percentage of correct predictions can be compared. Although war initiators tend disproportionately to be the stronger of two states, Bueno de Mesquita finds that the relationship between war initiation and states' expected utility for war is stronger than that between war initiation and relative capabilities. Moreover, the expected utility model is stable over time whereas the predictive power of the power politics model is considerably weaker in the twentieth century than in the nineteenth century. Though Bueno de Mesquita's cases are not restricted to situations involving deterrent threats, and though some might quarrel with his measurement of utilities solely in terms of alliance patterns, his findings provide further evidence against the proposition that one's military superiority is the primary determinant of the absence of aggression by the adversary. These and other findings suggest that deterrence practitioners should be at least as sensitive to the motivations of the potential aggressor as to the dyadic balance of military capabilities.

The Karsten, Howell and Allen Study

In Military Threats: A Systematic Historical Analysis of the Determinants of Success, Karsten, Howell and Allen attempt to test a wide range of hypotheses relating to the conditions contributing to the successful use of military threats.

---


27 Bueno de Mesquita, The War Trap, pp. 140–5. The argument that overall military capabilities have only a secondary impact on immediate deterrence would seem to be strengthened by the fact that most empirical studies of the relationship between the dyadic balance of military capabilities and war support the parity hypothesis rather than the power preponderance hypothesis (see fn. 13). The relationship between these studies, and our ability to make inferences, is complicated by the fact that they focus on slightly different sets of cases. The latter set of studies focuses on the direct relationship between pairs of states, whereas the Russett–Huth analyses focus on a more restricted set of cases involving the existence of (1) a prior threat and (2) one which is directed against a third party. Further research is clearly needed on the relationship between these different dimensions of deterrence.
They include over a hundred independent variables, but our focus here is restricted to those relating to the role of military capabilities. Karsten, Howell and Allen find that the balance of objective capabilities does not significantly affect the outcome of the crisis, and weaker or smaller states are no more likely than stronger states to yield to threats. But if the target perceives that the threatener has the capability to inflict serious damage (in absolute terms, regardless of the damage the target could inflict), threats tend to succeed. On the other hand, the threatener’s perceptions of the target’s capabilities tend to be far less important (although this has changed in the nuclear age). In fact, there is a modest tendency, at least in the pre-1945 period, for threats to fail when the threatener perceived itself as superior and to succeed when it perceived itself as inferior. This is explained in terms of the enormous importance of resolve and the underlying interests at stake, though the threatener’s perception of the target’s resolve is more important than the target’s perception of the target’s resolve. Karsten, Howell and Allen criticize the common focus on the credibility, resolve and signalling of the threatener and the common failure to recognize the importance of target goals and interests.

The validity of many of the Karsten, Howell and Allen findings is limited by some rather serious flaws in their research design. There is no clear theoretical framework guiding the study, and the cases are not systematically selected from any well-defined population, so that it is difficult to generalize beyond the specific cases included in the study. For a more sophisticated large-N study of the efficacy of military threats, and one which focuses explicitly on deterrent threats, we turn to the Russett–Huth research programme on extended deterrence.

THE RUSSETT AND HUTH STUDIES

Of all the empirical studies of deterrence, those by Russett and Huth are in many respects the most relevant for our purposes. They focus explicitly on deterrence

---


29 Karsten, Howell and Allen, *Military Threats*, pp. 54, 67–70, 73, 83, 110. The authors also conclude that neither the clarity of threats nor attempts to fine-tune threats have much impact on the outcome of the dispute. Whether clear and specific threats are more effective than more ambiguous and general threats is an important question. The Karsten, Howell and Allen findings draw some support from work on Leng’s Behavioral Correlates of War Project, which involves events data analyses of interstate crisis behaviour. Leng finds that the more specific the threat the greater the likelihood of a defiant response, for some degree of vagueness facilitates the target’s compliance with the threat by avoiding the appearance of being bullied. Russell J. Leng, ‘When Will They Ever Learn? Coercive Bargaining in Recurrent Crises’, *Journal of Conflict Resolution*, 27 (1983), 379–419, and ‘Reagan and the Russians: Crisis Bargaining Beliefs and the Historical Record’, *American Political Science Review*, 78 (1984), 338–55. Specific threats may also fail because the target is frequently able to ‘design around’ a deterrent threat, particularly if it has multiple options by which to achieve its objectives. See George and Smoke, *Deterrence in American Foreign Policy*, Chap. 17.
rather than on military threats in general, though they restrict themselves to the question of extended deterrence (deterrence of an attack against an ally) and do not examine the deterrence of a direct attack. Although their focus is on situational rather than decision-making variables, their most recent studies have begun to incorporate indictors of bargaining behaviour, and by supplementing their quantitative analyses with case studies they have begun to examine some intervening perceptual variables. Their ongoing research programme includes Russett’s ‘The Calculus of Deterrence’ (1963) and ‘Pearl Harbor: Deterrence Theory and Decision Theory’ (1967), Huth and Russett’s ‘What Makes Deterrence Work? Cases from 1900 to 1980’ (1984) and ‘After Deterrence Fails: Escalation to War?’ (1986), and Huth’s Extended Deterrence and the Prevention of War (1989). Although each study builds on the previous one it would be useful to consider each separately, for several reasons. The evolution of this research programme demonstrates the sensitivity of some empirical findings to the particular research design employed and also raises some important conceptual and methodological issues in the empirical study of deterrence.30

Russett, ‘The Calculus of Deterrence’

In his initial study of deterrence a quarter of a century ago, Russett frames his question of extended deterrence as follows: ‘How can a major power make credible an intent to defend a smaller ally from attack by another major power?’31 His objective is to identify the variables accounting for the success or failure of deterrent threats, and his method is a comparative study of all the cases during the last three decades where a major power ‘attacker’ overtly threatened a pawn with military force, and where the 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.

Russett generates seventeen such cases between 1935 and 1961. He then measures several independent variables, including the importance of the pawn (pawn/defender ratio of both population and GNP); the presence or absence of a prior formal commitment by the defender; the dyadic balance of both overall strategic and local military power; the nature of the defender’s political system (democratic or ‘totalitarian’); the extent of formal military co-operation between defender and pawn (arms transfers or military advisers); political interdependence between defender and pawn (defined generally in terms of a current or re-


cent alliance, recent occupation, close ideological ties, etc.); and economic interdependence between defender and pawn (relative proportion of imports and exports). The dependent variable is deterrence success or failure, with success defined as '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'.

The comparative analysis is conducted without the use of formal statistical methods. It is found that the effectiveness of deterrent threats is unaffected by the size of the pawn, the existence of a formal commitment by the defender, or the strategic or local balance of military capabilities, though equality on at least one military dimension is a necessary (but not sufficient) condition for a deterrence success. Deterrent threats by democratic regimes are slightly less credible than those of non-democratic regimes. Some level of military co-operation between defender and pawn appears to be a necessary condition for successful deterrence, but it is not sufficient. The existence of political ties is 'helpful if not essential', and economic interdependence is 'virtually essential' to successful deterrence. Russett then attempts to identify the factors associated with the defender's decision whether or not to go to war to defend the pawn once it has been attacked. Neither the size of the pawn, the military balance, nor the nature of the defender's regime has much of an impact on the defender's response, but bonds of economic, political and military interdependence are quite important (as they were for the attacker's actions). Russett gives great emphasis to these bonds between defender and pawn, and suggests that strengthening these ties is a means for the defender to increase the credibility of his deterrent threats. He fails to acknowledge, however, that increasing ties between defender and pawn to reinforce commitment and credibility may involve substantial costs.

Russett also presents an expected-utility model (though he does not refer to it in those terms) to explain the actions of both attacker and defender, though this model appears in footnotes and is not fully integrated into the analysis as a whole. The defender will pursue a 'firm' policy and attempt to deter the adversary if his prospective gains from successful deterrence, weighted by the probability of success and discounted by the cost and probability of war, exceed the losses of retreat (the failure to issue a deterrent threat). The adversary will attack in spite of the deterrent threat if (and only if) the expected value of attacking (as determined by the probability that the defender will not respond and the benefits of an uncountered attack and the cost and probability of war resulting from an attack which is resisted) exceeds the value of the status quo.

It would be useful to note some of the limitations of Russett's study, though some of these are corrected in subsequent work. First, Russett's theoretical discussion is weakened by framing the question in terms of the credibility of the threat. Credibility is an intervening perceptual variable which may help explain,
along with other variables, the effectiveness of the threat, but credibility is not equivalent to effectiveness. In fact, since Russett’s expected-utility model includes the value of a successful (i.e., uncountered) attack, it implies that if the value of the target is important enough to the attacker, the existence of a credible threat by the defender will not be sufficient to deter. But Russett’s analysis is not closely guided by the model he presents. He never directly tests whether or not the threat is believed (a perceptual variable), but only whether or not it leads to a successful outcome (a behavioural variable). This emphasis on the credibility of the threat affects the theoretical interpretation of the findings but not the validity of the observed empirical associations.\(^35\)

A more serious problem from the perspective of the validity and generalizability of the empirical findings is the definition of deterrence success to include cases in which an attack on the pawn is repulsed without violent military conflict between the two major powers.\(^36\) As a result, several cases that might normally be regarded as partial failures of deterrence are instead coded as instances of successful deterrence, including the Berlin Blockade (successful deterrence by the United States), Anglo-French attack on Egypt in 1956 (Soviet success), Chinese Communist artillery blockade of Quemoy in 1958 (US success), and Bay of Pigs (Soviet success).\(^37\) To the extent that the 1948 Berlin Crisis was a US success, for example, it was a success of compellence after a partial failure of deterrence. Similarly, the successful US threat against the Soviet Union in Iran in 1946 might be better classified as compellence than deterrence.

The more general theoretical problem is the treatment of the dispute outcome (success/failure) as a dichotomous variable, whereas George and Smoke argue persuasively that deterrence can fail in a variety of different ways and that the initiator can often ‘design around’ a deterrence threat. Though George and Smoke are undoubtedly correct on the theoretical level, I believe that for the purposes of a large-N correlational study the dichotomous classification of the dependent variable is a reasonable first approximation, but one which ought to be refined in subsequent research.

Perhaps an even more serious problem with defining deterrence success is that the very concept implies a causal relationship, that the potential aggressor does not attack the pawn because of the defender’s threat. The mere observation of non-attack is necessary but not sufficient to demonstrate the success of a deterrent threat. One must also show that in the absence of such a threat the adversary would have attacked the pawn, and that the adversary was dissuaded by the defender’s threat rather than something else (such as his domestic public opinion, anticipated diplomatic reaction, ability of the pawn to mount a successful or at least costly defence, etc.).

\(^35\) For further discussion of this problem, and a simple derivation from the expected utility model of the conditions under which a credible threat will be a necessary and sufficient condition for deterrence, see the critique by Clinton F. Fink, ‘More Calculations about Deterrence’, *Journal of the Conflict Resolution*, 9 (1965), 54–65.

\(^36\) Russett, ‘Calculus of Deterrence’, p. 98.

\(^37\) George and Smoke, *Deterrence in American Foreign Policy*, p. 517.
This counterfactual hypothesis regarding what the adversary would have done in the absence of the deterrent threat is, of course, difficult to demonstrate and involves substantial data requirements. Russett (and later Huth and Russett) attempts to deal with this problem by including the adversary's prior threat against the pawn as a definitional requirement for all cases, assuming such a threat to be a sufficient indicator of the adversary's intent to attack.\textsuperscript{38} Admittedly, this assumption is not always valid. An adversary may threaten a weaker ally not because it intends to attack, but instead as a means to some other end: to win a concession on some other issue, to distract the defender's attention, or to induce him to divert vital military resources to the defence of the pawn.\textsuperscript{39} Such a threat may also be a product of the domestic or bureaucratic politics within the adversary's regime. A case-study methodology can devote more attention to the motivations and intent of the potential attacker as a means of verifying the counterfactual, but even it can rarely be conclusive.\textsuperscript{40} In any case, such intensive analysis is not feasible for a large-\textit{N} study of this kind.\textsuperscript{41} My own view is that Russett's assumption is a good first approximation but that additional work to confirm the validity of this indicator of the adversary's intent would be very valuable.

In spite of its limitations, Russett's 1963 study marks the initiation of an important research programme. This programme is particularly impressive because of Russett's recognition of the costs of excluding perceptual variables and the limitations of correlational analysis for making causal inferences. In a 1970 article he discusses the need to use both correlational and case-study methods to complement each other for the purposes of the development of theory, and his case study of Pearl Harbor is consistent with this conception of theory building and cumulation in international research.\textsuperscript{42}

\textit{Russett, 'Pearl Harbor: Deterrence Theory and Decision Theory'}

One of the main purposes of the Pearl Harbor study was to explore and validate some of the empirical associations uncovered in the earlier correlational analysis. Russett's earlier expected-utility model is used as a framework for the

\textsuperscript{38} Russett, 'Calculus of Deterrence'; Huth and Russett, 'What Makes Deterrence Work' and 'After Deterrence Fails'; Huth, \textit{Deterrence and War}.

\textsuperscript{39} Fink, 'Calculations'.

\textsuperscript{40} This problem of identifying genuine deterrence success led George and Smoke to restrict their \textit{Deterrence} study to cases of deterrence failure. This is reasonable given their interest in constructing a typology of deterrence failure, but ultimately the success or failure of deterrence cannot be fully explained without a fully controlled study which includes cases of deterrence success as well as failure.

\textsuperscript{41} Admittedly, the \textit{N} of seventeen cases in Russett's 'Calculus of Deterrence' is too small to capture all of the advantages of a large-\textit{N} correlational study, but the number of cases is increased in his later studies.

theoretical interpretation of Japan’s decision for war. In spite of the absence of an explicit American deterrent threat, Russett argues, Japanese decision makers were fully convinced that the United States would respond militarily to a Japanese attack on Malaya and especially the Dutch East Indies because of the strategic and economic value of those colonies. This is consistent with Russett’s earlier argument that military, political and economic ties reinforce commitment in a way that formal threats do not.  

The fact that the Japanese perceived the US threat to be highly credible but attacked anyway demonstrates the fallacy (in Russett’s 1963 study) of equating the credibility of the threat with its effectiveness. Russett shows that the Japanese in fact perceived the military superiority of the United States and that the Japanese expected to lose a prolonged war if the United States chose to fight such a war. Russett explains the Japanese decision to attack in spite of the high military risks in part by their very low assessment of the value of the status quo, which was reinforced by domestic and bureaucratic political as well as strategic and economic considerations.  

Thus Russett rejects the common argument that the Japanese attack on Pearl Harbor was an irrational action, and argues that given Japanese preferences, expectations and constraints, their behaviour was consistent with a rational expected-utility calculus. He acknowledges, however, that two aspects of Japanese behaviour were not consistent with a rational model: the assumption that the United States would prefer to reach a negotiated settlement recognizing Japanese hegemony in Southeast Asia than to fight a prolonged war, and the failure to explore the validity of that assumption. On the basis of this case Russett argues that a theory of extended deterrence must include in the set of possible outcomes a direct attack against the defender as well as inaction and an attack against the pawn.  


This article develops the expected-utility model from Russett’s ‘Calculus of Deterrence’ piece and tests the model on cases of extended, immediate deterrence between 1900 and 1980. It makes a number of improvements on Russett’s first study. It recognizes the problem of inferring the success of deterrence and argues that by explicitly focusing on immediate deterrence the problem is reduced though not eliminated. By extending the temporal domain of the study

---

43 Russett, ‘Pearl Harbor’, pp.94–6. This argument is reminiscent of Jervis’s (1970) distinction between signals and indices.
45 Russett classifies this case as a deterrent failure, but in fact it involved the backfiring of a strategy of compellence as much as the failure of deterrence. The United States attempted not only to deter the Japanese from moving into Southeast Asia but also to compel them to withdraw from China, and used a highly coercive oil embargo to enforce that policy.
46 In the nuclear age the option of a direct attack against a major power defender rather than its pawn is unlikely to be very attractive.
back to 1900 and forward to 1980, it significantly increases the number of cases to fifty-four. This permits the use of more formal statistical methods, which were not used in Russett’s 1963 study. Deterrence failure occurs if there is an attack against the protégé (in a semantic change, the authors refer to the ‘protégé’ rather than the ‘pawn’). Such an attack is operationally defined as ‘a government-sanctioned engagement of its regular armed forces in combat with the regular armed forces of the protégé and/or its defender, resulting in more than 250 fatalities’. Cases in which the attacker gains its principal goals or occupies the territory of the protégé in spite of minimal fatalities are also classified as failures of deterrence. All other cases are classified as deterrence successes. Fifty-seven per cent of all cases are classified as successes.47

This study also introduces some new independent variables and refines some of the operational indicators for variables used in earlier studies. Hypotheses based on five different dimensions of relative military capabilities are now tested: overall military and economic capabilities, or military potential; overall existing military capabilities; potential local capabilities proximate to the protégé; existing local capabilities; and defender’s possession or non-possession of nuclear weapons. Economic and military capabilities are measured using the Correlates of War capability data, with a loss-of-strength gradient introduced for local capabilities.48 The effects of the defender’s previous behaviour is introduced through dummy variables tapping whether or not the defender had fought on behalf of the protégé in the previous deterrence situation (against any adversary). The presence or absence of a formal military alliance between defender and protégé is also included. Economic ties are measured by the protégé’s share of the defender’s total merchandise exports and imports, and political–military ties are measured by a four-point scale of the percentage share of major weapons systems imported by the protégé from the defender. The strategic importance of the protégé to the defender is measured (as before) by the ratio of the protégé’s capabilities (on the various dimensions) to those of the defender, and whether or not they shared a common border. Although the Pearl Harbor study demonstrated the critical importance of the expected utility of peace to the attacker, this variable is very difficult to measure and is not incorporated into the model, which is a serious limitation and one acknowledged by the authors.49 Each of the above factors is an independent variable in a linear model predicting the success or failure of deterrence. The relationships are analysed by the technique of probit analysis, which is similar to regression analysis but more appropriate in the case of a dichotomous dependent variable.

The model predicts 78 per cent of all outcomes correctly. The variables having a significant ($p < 0.10$) impact on deterrence success or failure include economic and political ties (trade and arms transfers) and the local military balance. The existence of a formal military alliance (formed prior) to the aggressor’s initial

48 Singer, ‘Material Capabilities’.
theflight)betweendefenderandprotégéismoderatelyimportant,buttendsto
reduce the likelihood of successful deterrence. The overall strategic balance
and the defender's behaviour in previous crises, however, appear to have little
impact on the effectiveness of deterrence. Moreover, the defender's possession
of nuclear weapons has only a 'marginal' impact on outcomes. This leads Huth and
Russett to conclude that 'local military capabilities (of the defender and protégé
combined) seem to have more to do with successful deterrence than do strategic
capabilities, and both may be less important than having a dense network of
political and economic bonds between defender and protégé.' This is consistent
with the findings of other studies surveyed above that the balance of military
capabilities is not the primary determinant of deterrence.

After examining the sources of deterrence success and failure, Huth and
Russett analyse the factors determining whether the defender will fight to defend
its protégé, which happened in fifteen of their twenty-three cases. The defender is
more likely to fight if the protégé is important in terms of its relative military
capability, and particularly if there exists a military alliance between defender
and protégé. However, the dyadic military balance between defender and
attacker, geographical contiguity, and the defender's past behaviour are not
associated with the occurrence or non-occurrence of war.

It is important to note that the variables associated with deterrence success
and those associated with the defender's decision to fight if deterrence fails are
not the same. The only factor important in both decisions is the existence of a
military alliance between defender and protégé, but the effect of alliances re-
verses the direction in the second decision. This is a particularly dangerous com-
bination, and leads Huth and Russett to conclude, again emphasizing the
importance of economic and political ties between defender and protégé, that 'a
military alliance not backed up by more tangible linkages ... may increase the
danger that a defender will be drawn into war'. The tendency for alliances to
courage the adversary to attack the protégé and then to lead the defender to
intervene would appear to imply a mismatch between the calculations of the

50 Huth and Russett explain this by arguing that if an adversary threatens a protégé that is form-
ally allied to the defender, that adversary has already decided to stand firm and carry out the threat

51 Huth and Russett, 'What Makes Deterrence Work?', pp. 516–18. Similarly, Blechman and
Kaplan, Force without War, find no evidence that US decisions to use force short of war are strongly
influenced by aggregate strategic capabilities, and argue that the local balance of conventional
power tends to be more important. Organski and Kugler reach similar conclusions from their exami-
nation of fourteen cases of deterrence since 1945 (The War Ledger, Chap. 4). They find that nuclear
powers have prevailed in only about half of these but that conventional superiority does make a dif-
ference. These findings are reinforced in a subsequent study, where Kugler finds that even a nuclear
monopoly has brought a favourable outcome only about half the time (Jacek Kugler, 'Terror without
Deterrence: Reassessing the Role of Nuclear Deterrence', Journal of Conflict Resolution, 28
(1984), 470–506). Weede finds that mutual nuclear deterrence has reduced the risk of war not only
between superpowers but also between allies in opposing blocs (Erich Weede, 'Extended Deterrence

attacker and the defender, but Huth and Russett suggest an alternative interpretation in which economic and political ties are the primary influence on the attacker’s decision and alliances are more important for the defender’s decision.\textsuperscript{53}  


This paper, like the previous one, deals with both the determinants of the success of immediate extended deterrence and with the determinants of escalation to war in the event that deterrence fails. It goes beyond the previous study by suggesting additional hypotheses, introducing new variables and conceptualizing others, modifying some definitional criteria and empirical indicators, extending the temporal domain and by making the corresponding changes in the database. As a result, there are some new empirical findings.

More specifically, the temporal domain of the study is extended back until 1885. This, plus further additions and deletions based on the examination of previously unused sources, results in an increase in the number of cases from fifty-four to fifty-eight.\textsuperscript{54} In one particularly significant change from the 1984 study, the measure of the past behaviour of the defender (whether or not it came to the defence of the protégé) is altered to tap its behaviour in the last deterrent situation with the same adversary rather than with any adversary. The concept of relative military capabilities (and hypotheses related to them) is also modified and improved, and now includes immediate, short-term, and long-term dimensions, plus the defender’s possession/non-possession of nuclear weapons. The immediate balance, measured by forces currently and proximately available to attack or defend the protégé, are hypothesized to be the most important for both deterrence success and for the defender’s decision whether to intervene in the case of deterrence failure. The short-term balance consists of all active-duty forces and readily available reserves. The long-term military balance is defined as all existing military forces and national mobilization capabilities (economic, industrial, demographic) for fighting a protracted war.\textsuperscript{55} It is hypothesized that

\textsuperscript{53} Huth and Russett, ‘What Makes Deterrence Work?’, pp. 522–3. They hypothesize, in the ellipses in the quoted passage, that these linkages ‘perhaps give the defender some control to prevent adventurism by a protégé’. Alternative explanations would have to be considered, however, including one based on selection bias, a point to which I will return. I should note that the statement quoted in the text is not technically supported by their analysis. It implies an interaction effect between military alliances and tangible (i.e., economic or political) linkages, but no such term is formally incorporated into their model.

\textsuperscript{54} Some cases were deleted because of the ambiguity regarding whether a prior threat of attack actually existed (and thus whether this was actually a case of immediate deterrence). Other cases were deleted because the intervention of the defender occurred after hostilities (between attacker and protégé) had already escalated to the level of large-scale armed conflict, leading to the classification of the case as compellence rather than deterrence. See Huth, \textit{Deterrence and War}.

\textsuperscript{55} The immediate balance corresponds to the ‘existing local’ balance in the 1984 study; the short-term balance to the ‘existing overall’ balance; and the long-term balance to the ‘potential overall’ balance.
the attacker does not usually expect a long war of attrition and therefore discounts the importance of the long-term balance of capabilities, and that he may attack in spite of his inferiority in long-term forces as long as he has superiority in immediate or even short-term forces. The long-term balance is hypothesized to be important only for the defender's decision regarding whether to intervene to save the protégé after the failure of deterrence, but even that is qualified.

The most significant theoretical change from the earlier Russett/Huth studies is the inclusion of the bargaining behaviour of the adversaries into the model. The basic hypothesis is that the effectiveness of diplomatic and military techniques used by the adversaries to influence each other is a non-linear function of one's bargaining behaviour, with threats of moderate firmness being the most effective. Excessively conciliatory bargaining behaviour undermines credibility and therefore deterrence, and excessively hostile threats provoke equally hostile responses and trigger an upward spiral of escalation.\(^{56}\)

The definition of deterrence success or failure is the same as for the 1984 study. Failure is defined as 'an attack on the protégé by regular military forces resulting in more than 250 fatalities, where the attacker gained its principal political or territorial goals even though fatalities were minimal, or where the attacker occupied territory of the protégé for several years'. Success is the absence of such an attack. This implies that an attack which is repulsed with minimal casualties by the protégé or the deterrer is defined as a deterrent success. Thus the 1948 Berlin case is still classified as a deterrent success for the United States, which illustrates the problem of failing to include a category for the partial failure of deterrence. Another limitation of the 1984 study which is not corrected here is the failure to incorporate the attacker's evaluation of the acceptability of the status quo, the importance of which was demonstrated by Russett's case study on Pearl Harbor.\(^{57}\)

\(^{56}\) Drawing on some of Leng's early work on the Behavioral Correlates of War project, Huth and Russett have coded the diplomatic and military actions of both defender and attacker for each case of attempted deterrence. Diplomatic strategies are classified as bullying, conciliatory or firm-butfair. Military actions are classified as either policies of tit-for-tat, strength or weakness, depending on whether they matched, exceeded or failed to match the attacker's level of escalation at each stage in the crisis. These many pairs of actions over the course of the crisis are then aggregated into a single measure reflecting the 'predominant influence strategy' of a state on both diplomatic and military dimensions. For a discussion of these classification systems see Russell J. Leng and Hugh G. Wheeler, 'Influence Strategies, Success and War', *Journal of Conflict Resolution*, 23 (1979), 655–84.

\(^{57}\) Huth and Russett, 'After Deterrence Fails'. Some aspects of the initiator's evaluation of the status quo may be incorporated indirectly through bargaining behaviour and reputation. Bullying strategies are hypothesized to be ineffective precisely because of the diplomatic and domestic political costs of retreating from the status quo. But a loss incurred from a retreat from the status quo is analytically distinct from the value of the status quo itself, so I would argue that the value of the status quo has not been incorporated.

This raises two very interesting theoretical questions. One is whether states evaluate prospective gains and losses in terms of the value or utility of the outcome or final asset position, as classical microeconomic utility theory assumes; or whether calculations are based on the magnitude and direction of the change in utility, as prospect theory assumes. (The more appropriate question would be the relative weights given to final asset positions and to changes in assets or utilities.) A related
Applying probit analysis to the data, Huth and Russett find that successful deterrence is associated with an immediate or short-term balance of forces favouring the defender ($p < 0.05$), with the long-term balance being only weakly associated with the outcome. The defender’s possession of nuclear weapons, or an overt threat to use them, has no impact on the outcome of the crisis. In the most surprising finding, and one contrary to the central result of the 1963 and 1984 studies, economic and political–military ties between defender and protégé are found to be unrelated to the success or failure of deterrence. Unfortunately, the authors make little effort to explain this dramatic change from their previous studies, though they do assert that it derives from the introduction of crisis bargaining variables into the analysis. These bargaining variables are found to be significantly correlated with the success or failure of deterrence. As hypothesized, reciprocal strategies are found to be associated with successful outcomes. Deterrence is likely to succeed if the defender follows a firm-but-fair diplomatic strategy, and likely to fail for defenders following conciliatory or bullying strategies ($p < 0.01$). Similarly, tit-for-tat policies of military actions are usually successful, whereas excessively firm or cautious policies are not ($p < 0.025$).

Another finding that runs contrary to the results of the 1984 study is that the outcome of a crisis is associated with the reputation of the defender, defined here as its behaviour in the last crisis against the same adversary. Both backing question (the importance of which would be rejected by expected utility theory) is whether the initiator, having made the threat, conceives of a prospective retreat as a retreat from the status quo or a retreat to the status quo. That is, would a retreat be seen as a loss or the absence of a gain. This question of how the decision is framed may be critical. There is substantial evidence in social psychology that more weight is given to losses than to gains and that individuals are risk-acceptant when faced with losses and risk-averse when faced with gains. See Daniel Kahneman and Amos Tversky, ‘Prospect Theory: An Analysis of Decision under Risk’, *Econometrica*, 47 (1979), 273–91.

58 Huth and Russett, ‘After Deterrence Fails’. Only the statistically significant (at $p = 0.10$) probit coefficients are reported, so I must rely on the authors’ interpretation of the results.

59 The authors acknowledge that the generalizability of this finding is restricted by the limited number of cases involved (eighteen with nuclear defenders, four with overt nuclear threats).

60 One possible explanation is that the defender’s bargaining strategy is a function of the extent of defender–protégé bonds (i.e., the stronger the ties the more coercive the bargaining), with the resulting multicollinearity accounting for the drop in significance of defender–protégé bonds. The correlation between these two variables is not reported in the article, though I have been informed (Huth, private correspondence) that correlations between bargaining behaviour and defender–protégé bonds (economic ties and arms transfers) are very low (0.08 and 0.15, respectively), suggesting that multicollinearity is not a problem.

61 This finding of the effectiveness of reciprocal strategies is consistent with a growing body of laboratory studies. See Robert Axelrod, *The Evolution of Cooperation* (New York: Basic Books, 1984) and numerous studies by other game theorists and social psychologists. It is also consistent with the findings of Leng and his colleagues on the Behavioral Correlates of War project. Leng and Wheeler (‘Influence Strategies’) find, in their data-based analysis of twenty crises since 1900, that reciprocating strategies are the most successful, particularly against an adversary employing a bullying strategy. Bullying strategies tend to generate coercive counter-threats and escalation rather than compliance. See also Leng, ‘Reagan and the Russians’.

62 In ‘What Makes Deterrence Work?’ Huth and Russett defined the defender’s reputation in terms of its behaviour in the previous crisis with any adversary, and found that it has no significant impact on the outcome of a current crisis.
down and prevailing in the last crisis against the same adversary are associated with the failure of deterrence, which leads Huth and Russett to conclude that 'a stalemate is a safer outcome'. This conclusion, however, must be regarded with a certain amount of scepticism in the absence of further research. The sensitivity of many of these empirical findings to relatively minor changes in the operationalizations of certain variables or to the addition of new variables is one ground for caution. In addition, the failure to examine alternative explanations for the selection of different bargaining strategies raises the possibility of spurious inferences regarding the causal impact of these strategies.

Huth and Russett conclude that the attacker's decision appears to be based primarily on short-term considerations, particularly military ones, and that long-term military power avails little. What is important is the defender's ability to prevent a quick seizure of territory or to roll back that seizure relatively promptly, not the threat of future retaliation. They infer from this that deterrence by denial is more critical than deterrence by punishment for immediate extended deterrence.

Now let us consider the conditions affecting the defender's decision to go to war in the event that deterrence fails. The model correctly postdicts 83 per cent of these cases \((n = 24)\). The short-term balance of military forces in being is statistically significant, and the long-term balance is nearly significant. The defender's ties with the protégé are also important, particularly formal alliances and geographical proximity. Crisis bargaining behaviour is statistically significant \((p = 0.10)\).

Thus several factors which appear to be important in the attacker's decision whether or not to defy a deterrent threat are relatively unimportant or less im-

---

63 Leng finds that while reciprocating strategies tend to work best, states tend to adopt more coercive bargaining strategies in successive crises against the same adversary. The diplomatic victor in one dispute tends to utilize the same (successful) influence strategy in the next crisis with the same adversary, unless the adversary adopted a more coercive strategy, in which case (and only under such conditions) the previous winner would also adopt a more coercive strategy. The loser, assuming that its diplomatic defeat was due to the failure to demonstrate sufficient resolve, tends to adopt a more coercive strategy in the next crisis. A diplomatic compromise tends to result in more coercive influence strategies by both parties in the next crisis. Crises ending in war tend to result in more coercive strategies in the next crisis unless a state perceives that the war had been 'unwanted' (i.e., one's behaviour was overly coercive, leading the adversary to preempt), in which case a more accommodative strategy is adopted. See Leng, 'When Will They Ever Learn?'

64 This finding raises another question. If 'stalemate' is defined as a possible outcome of previous crises, utilized in the statistical analysis, and found to be associated with successful deterrence in a current crisis, that is not made explicit in this study. And if outcomes can be adequately measured trichotomously for use as an independent variable predicting to behaviour and outcomes in the next crisis, this new measure should be utilized as the dependent variable in all crises, replacing the problematic success/failure dichotomy. A technical methodological point is also in order. Statistical inference generally requires the independence of cases. If the outcome of one case is affected by the outcome of the previous case, the assumption of independence is violated, and more sophisticated statistical procedures are normally required.

important in the defender’s decision whether to fight in the event that deterrence fails. Whereas the attacker is influenced by immediate and short-term military considerations and by its adversary’s reputation and bargaining strategy, and basically undeterred by political–military and economic ties between defender and protégé, the defender is more influenced by long-term military considerations and defender–protégé bonds. It is less influenced by its own past behaviour, by the nature of the bargaining process or by short-term military considerations. The fact that the decisions of the attacker and defender are influenced by such different criteria increases their insensitivity to the cost–benefit calculations made by the other and thus increases the likelihood of a conflict spiral driven by misperceptions. States are particularly likely to underestimate the costs of retreating to the adversary, and the failure of the attacker to consider the value to the defender of economic and political–military ties between defender and protégé is potentially a critical factor contributing to war.66

B/T CAPABILITIES: THE PROBLEM OF SELECTION BIAS

One of the themes emerging from most of the studies surveyed above is that the possession of military superiority by the defender is no guarantee that deterrence will work. The overall balance of military capabilities between two states has at most a secondary impact on the success or failure of immediate deterrence (and on the effectiveness of military threats in general). Moreover, although strategic deterrence may be effective in regulating the behaviour of nuclear states towards each other, there is little evidence that nuclear weapons affect the success or failure of extended deterrence of attacks on allies.

It is essential to recognize, however, that there is a serious but perhaps unavoidable flaw in these research designs regarding the role of military capabilities in immediate deterrence, and that it is difficult to assess the extent to which this biases the empirical findings. In order to lend plausibility to the inference that the absence of a military attack constitutes an indicator of successful deterrence, there needs to be evidence that the potential attacker would have attacked in the absence of a deterrent threat. Although this can never be conclusively demonstrated, at a minimum there is a need for some evidence that he was seriously considering military action. Several studies fail to provide such evidence, and this limits the relevance of their results for the question of immediate deterrence. The Huth and Russett studies attempt to deal with this problem by requiring the existence of the aggressor’s prior threat against the protégé as a definitional requirement for a deterrence situation, and assume such a threat implies that the attacker intended to act militarily.67 Although some assumption of this kind may be necessary, it may result in a biased selection of cases included in the study and consequently an underestimation of the importance of military capabilities in immediate deterrence.

66 Huth and Russett, ‘After Deterrence Fails’.
67 Earlier I noted the limitations of such an assumption.
The problem is that a research design incorporating the assumption of a prior threat permits an analysis of the conditions for the success or failure of immediate deterrence but not of the conditions under which the attacker makes the threat that initiates the crisis and thus qualifies the case for inclusion. If the attacker makes threats only in those instances in which its expected probability of success (as determined by the dyadic military balance) is sufficiently high, then because of selection bias the importance of capabilities may not be reflected in an analysis of the attacker's later decision whether or not to defy the deterrence threat and attack. If threats are made primarily by 'stronger' states (defined in terms of their military strength) and then implemented if and only if their immediately and proximately available military forces are expected to be sufficient, the empirical results might not be much different from those actually observed. This very plausible alternative explanation cannot be ruled out by Huth and Russett or by others studying disputes involving a prior threat.

Note that a similar problem would arise in an analysis of the value of the protégé to the attacker in immediate deterrence. To the extent that the initial threat is made only against important targets, selection bias will result in an underestimation of the impact of this variable in immediate deterrence. Combining the capability (which affects the expected probability of success) and value variables, we see that the problem of selection bias precludes a straightforward analysis of the role of the initiator's expected utility of attack in immediate deterrence.

The potential seriousness of the problem of selection bias is demonstrated by the finding that deterrence is significantly more likely to fail if the protégé has a pre-existing military alliance with its major power defender. As Huth and Russett recognize, the very fact that the aggressor makes a threat under such

---

68 Because these studies deal only with those cases in which threats of force have already been made, general deterrence has already failed, but that failure does not necessarily imply the failure of immediate deterrence. Thus the conditions for the success of immediate deterrence are not necessarily the same as those for the success of general deterrence, and vice versa. Overall military capabilities may be more important for general deterrence than for immediate deterrence.

69 This methodological problem is complicated further by the multiple dimensions of military capabilities and the dyadic military balance. Selection bias occurs only if the same dimension of military strength (or any other variable) affects both the initiation of the threat (and thus the case selection) and the decision whether or not to defy the defender's deterrent threat. Selection bias would not arise, for example, if the initial threat were made on the basis of one's overall military strength and if the actual attack were made on the basis of the initiator's expectation of a fait accompli based on immediately available military forces. There are good reasons to believe, in fact, that the factors influencing behaviour during a crisis may be different from the factors determining behaviour in the pre-crisis period. Domestic political pressures, vested bureaucratic interests and ego investment deriving from the prior decision to initiate a threat, reputational considerations, and increased propensities to take risks when faced with the alternative of losses from backing down, as well as the motivated biases generated by these considerations, all change the cost–benefit calculus for decision makers once a crisis has arisen.

70 As noted, Huth and Russett do not include this variable in either 'What Makes Deterrence Work?' or 'After Deterrence Fails'.

71 Another clear example of likely selection bias is the finding by Karsten, Howell and Allen (Military Threats, p. 83) that threats tend to fail when the threatener sees itself as stronger and succeed when it perceives itself as the weaker party.
conditions suggests its determination to carry out the threat regardless of any attempt by the protégé’s major defender to deter that action. This suggests that the critical variables affecting both the decision to issue the initial threat and the decision to attack are the interests and motivations of the aggressor rather than the presence or absence of alliances, and that the failure to incorporate these variables into the analysis may result in spurious inferences.72

Both the Karsten, Howell and Allen and Huth and Russett studies acknowledge this general problem. The latter caution that these findings apply only to cases of immediate deterrence . . . in which an overt military threat from a potential attacker has already become manifest. Perhaps this kind of case understates the value of military strength in general deterrence. If military power were overwhelming, possibly no aggressor would ever rise to the level of making the overt challenge that characterizes these cases.73

This selection bias affects not only the analysis of general deterrence but also the interpretation of the results in cases of immediate deterrence, because spurious inferences from observed correlations cannot be ruled out. This problem concerns any variable which affects both the criteria upon which the case selection is based and the success or failure of immediate deterrence. We are caught in a difficult dilemma. If we make the distinction between general and immediate deterrence, and examine only those cases in which there is some indication of a prior intention to attack, we risk introducing serious selection biases into the analysis. If, on the other hand, we make no distinction between general and immediate deterrence, we seriously diminish our ability to make the causal inference that the absence of an attack is due to successful deterrence. Although Huth and Russett not only acknowledge this dilemma but also deal with it in a reasonable manner, there may be alternative methodologies for dealing with the problem of selection bias, and the exploration of these constitutes an important avenue for further research. It would be useful to supplement the analysis of immediate deterrence with a separate analysis of the conditions under which military threats are initially made, or, more generally, of the conditions under which one state seriously considers an attack against another. The application of some sophisticated new statistical methods developed for quasi-experimental research designs utilizing non-random samples would be one possible approach.74 Case-study methods might also be very useful here, for the intensive

72 Huth and Russett, ‘After Deterrence Fails’, p. 17. One possible way to circumvent this problem would be to analyse those cases in which alliances are formed after the initial threat but before an attack. This would not fully deal with the problem of selection bias, however, because the alliance formation would probably be the causal result of the protégé’s (or defender’s) anticipation of an attack (based on the initiator’s prior threat). For a more general discussion of the causal relationship between alliances and war, see Jack S. Levy, ‘Alliance Formation and War Behavior: An Analysis of the Great Powers, 1495–1975’, Journal of Conflict Resolution, 25 (1981), 581–614.


analyses of individual cases may be particularly useful in determining the presumed aggressor's prior motivations and intent regarding the possible use of force. Statistical and case-study methodologies might be usefully combined, with case studies of well-selected cases used as validity checks for the operational indicators used in aggregate studies.75

The analysis of the conditions under which threats are initially made takes us back, of course, to the question of general deterrence. Although the conceptual distinction between general and immediate deterrence is useful for many purposes, it raises some difficult analytical problems. Thus, still another task for future research is to reassess the utility of the distinction between general and immediate deterrence.

CONCLUSIONS

The quantitative empirical studies surveyed above deal with different aspects of the question of when deterrence works and identify different independent variables. They also utilize different operational indicators, temporal domains, types of data and methods of analysis. Yet several general themes do emerge. One concerns the role of military capabilities in deterrence. We have seen that the overall balance of military capabilities between two states has at most a secondary impact on the success or failure of immediate deterrence (and on the effectiveness of military threats in general), though this finding must be regarded as tentative until the magnitude of selection biases can be assessed. Several studies suggest, however, that the balance of conventional capabilities in proximity to the target does play an important role in extended deterrence. Taken together, these findings are consistent with a more general theme emphasized by George and Smoke and by George in his work on crisis management: the utility of military threats, whether deterrent or compelling in nature, depends on the threatener's possession of a spectrum of military capabilities and options appropriate to the level of the threat and the behaviour it is attempting to influence.76

The finding that superior military capabilities alone are not necessarily sufficient for deterrence can be explained in part by the tremendous importance of the interests and resolve of the initiator of the crisis, which is another theme emerging from several of these studies. Bueno de Mesquita demonstrates the explanatory power of an expected utility model incorporating the initiator's expected benefits from war as well as its expected probability of success. Maoz and Karsten et al. also give greater emphasis to the motivations and behaviour of the initiator than to the capabilities, commitment and signalling of the defender.77


76 George and Smoke, Deterrence in American Foreign Policy; George, 'Crisis Management: The Interaction of Political and Military Considerations', Survival (1984), 323–34.

77 Bueno de Mesquita, The War Trap; Maoz, 'Resolve'; Karsten, Howell and Allen, Military Threats; Russett, 'Pearl Harbor'.
These factors are not easy to incorporate into large-\( N \) aggregate studies, however, for their measurement often requires a fairly intensive examination of individual cases. The case-study literature has been more successful than the quantitative empirical literature in dealing with perceptions, motivations and preferences of the actors. It has also been more successful than the quantitative literature in analysing the domestic incentives and pressures which often lead states to undertake military action in spite of the existence of credible threats by the defender.\(^78\)

Another important theme emerging from several of these studies concerns the stabilizing effects of reciprocity in the interaction between adversaries.\(^79\) These are important findings, particularly since they are reinforced by other social scientific research on reciprocity.\(^80\) Although these studies are reasonably rigorous and systematic, the operationalization of reciprocity in large-\( N \) studies is a difficult task, and this is a sufficiently important theoretical question to require further research from several different theoretical and methodological perspectives. Another aspect of deterrence theory requiring further research is the role of reputation, or the impact of the past behaviour of the adversaries. This variable has received little attention in the quantitative empirical literature on deterrence, not only because of the difficulty of constructing an operational indicator for a large-\( N \) study but also because of the limited development of this variable in the theoretical literature. One particularly interesting but highly tentative finding emerging from the Huth and Russett studies is that the success or failure of deterrence is associated with the defender’s response in the previous deterrence situation with the same adversary but not with its behaviour against another adversary.\(^81\) If confirmed, this would suggest that some recent theorizing which treats reputation as a one-dimensional concept may be misleading.

This review of quantitative empirical studies of deterrence has demonstrated that this is not an isolated body of literature, but instead has substantial intellectual interconnections with both rational deterrence theory and with the case-study literature on deterrence. Many of the hypotheses which it has systematically tested were derived from these other bodies of literature, and the quantitative empirical literature has in turn contributed to the development of theoretical generalizations regarding when deterrence works. New developments in both the formal and non-formal theoretical literature on deterrence provide a rich source of important hypotheses which should now be tested


\(^80\) See Axelrod, *The Evolution of Cooperation*.

\(^81\) Huth and Russett, ‘What Makes Deterrence Work?’ and ‘After Deterrence Fails’. 
empirically by systematic quantitative analysis as well as by case-study methods. In addition, some of the findings from quantitative empirical studies represent interesting theoretical anomalies and ought to stimulate additional theoretical analysis. The broader literature on deterrence provides a useful model of how formal theoretical, quantitative empirical and case-study methodologies can supplement each other to enhance our understanding of important theoretical questions with undeniable implications for contemporary international security.