

CHAPTER 1

Introduction

Great quarrels . . . arise from small occasions but seldom from small causes.

—Winston Churchill

Perfection of means and confusion of goals seem, in my opinion, to characterize our age.

—Albert Einstein

For several decades during the Cold War, the “unrestricted globalism” of containment policy left a strong imprint on U.S. strategic thought.¹ The question of what, if any, national interests the United States, or any other power for that matter, might have had in different parts of the world was not generally raised. Instead, many scholars and policy-makers were puzzled over two issues regarding effective deterrence.² The first of these issues concerned the development of the capabilities that would provide an effective deterrent and, if deterrence failed, give the capacity to carry out a threat. Central to this problem was the question of whether the strategic parity between two superpowers, reinforced by the threat of mutually assured destruction (MAD), was sufficient enough to deter the escalation of a minor dispute into a nuclear confrontation. Although *prima facie* disconnected, this pivotal concern of nuclear strategic thinkers was perfectly consistent with the research of a number of scholars in the general literature of international relations. Their similar focus on relative capabilities was largely shaped by the realist paradigm, which considered power as the omnipresent factor in world politics.

The second predominant issue in strategic theory and policy resulted from the belief that the failure to stand firm against an opponent in one region would reflect negatively on the U.S. reputation to meet threats in other parts of the world. Consequently, strategic thought was mainly directed toward developing the techniques of signaling U.S. intention to use force, even when its stakes were low in the

particular area of dispute. In the post–Cold War era, however, the demise of a major global adversary has prompted the question of whether U.S. national interests and goals should be uniform across all world regions. Strategic theories, mainly shaped in the context of superpower deterrence, do not always have well-developed answers to this important policy question.

To understand the intricate nature of credible threats in the nuclear age, the deterrence literature has thus centered almost exclusively on the problems of relative power and manipulative strategies. Astonishingly, it has almost completely bypassed the key question of what creates a genuinely credible reason to want to use force in the first place. What issues are at stake? How strong do the stakes, that is, a deterrer's national interests, have to be for an opponent to believe the deterrer will carry out the threat? Is the belief about the deterrer's credibility predicated on the perceptions of the deterrer's inherent interests, regardless of how costly the intentions are communicated? The question of what, if any, political goals a deterrer and potential attacker have in the dispute over a third nation—the paradigmatic situation between major powers, and the one where the credibility of threats is most questioned—has been largely ignored.

This book takes the position, rarely analyzed elsewhere, that regional stakes are a main ingredient of the credibility of deterrence between major powers. Specifically, the salience of a particular region for a major power's national interests indicates the inherent credibility of its (Defender's) threat to retaliate if another major power (Challenger) attacks the third nation located in such a region. The regional stakes refer to the various forms of linkages between the Defender and the area in which the third nation (Protégé) is located, such as politico-military, diplomatic, and alliance ties, or the degree of economic interdependence measured by a high volume of trade and investments. Besides issues directly related to their own territory, I believe that it is regional ties that precisely provide the key cues to the national interests of major powers.

Since “the criterion of ‘national interest’ should assist the policymaker to cut through the complex, multivalued nature of foreign policy issues” (George 1980, 234), it is one of the goals here to clarify the concept of threat credibility in terms of national interests. My main argument is that the national interest, shaping the inherent credibility of threats, sets the limits to the impact of the other two factors on the opponent, i.e., a deterrer's capacity to carry out its threat and manipulative strategies it uses to communicate a strong resolve. It is national

interest—in other words, the involved stake in the dispute—that shapes the opponent's perception of the deterrer's resolve. If there is an asymmetry of stakes to a potential attacker's advantage, for instance, the deterrer's persuasive power to retaliate if another nation is attacked is significantly diminished. On the other hand, if the asymmetry of stakes is to the deterrer's advantage, the maintenance of overkill capacity or the use of various commitment strategies to defend an ally could be quite costly and, more important, superfluous.

Why assume that an opponent is oblivious of a deterrer's *vital* national interests? It could be the case, but only if the deterrer has not defined its vital interests and prioritized its foreign policy goals. "Of course in many cases the balance of interest is ambiguous. . . . But at minimum, states seem able to separate peripheral from vital interests" (Jervis 1991, 42). Remarkably, this simple, but fundamental logic of foreign policy calculations, nuclear or nonnuclear alike, driven by differentiation between vital and peripheral interests, has been overshadowed by other concerns, as discussed previously. The main point is that simple power calculations and building a reputation for strong resolve should follow rather than precede the issue of foreign policy goals and the priority of stakes, which, in turn, shape inherently credible threats in deterrence situations.

The occasional confusion about foreign policy goals seems then to be yet another legacy of the Cold War era. Phrases such as "conceptual poverty," "from containment to confusion," or "paradigm lost" are consequently found as descriptors of U.S. foreign policy in recent critical literature (e.g., Haas 1995; Posen and Ross 1997; Nye 1999). An emerging number of critics point to the problematic nature of foreign interventions based on reputational reasons, instead advocating the need to formulate a "grand strategy" as the guiding principle for post-Cold War policy. It is the strategy that "would be based on the belief that concrete vital interests should determine U.S. commitments (rather than credibility determining commitments and commitments, in turn, determining interests)" (Layne 1997, 115; see also Johnson 1994, 144; Haas 1995; Nye 1999).

Clearly, current world events require a fresher look at and a reinterpretation of a number of old premises inherited from the past. This book points to one of the ways to rethink current and past events as they have emerged in relations between major powers. It is closely related to recent literature that emphasizes the need for grand strategy as opposed to ad hoc micromanagement of foreign policy crises. In order to make such a transition in the post-Cold War era, I argue that

it is essential to bring back into focus the long-neglected question of the role of national interests in both deterrence theory and strategic policy. This argument is strongly supported in my quantitative analysis of deterrence crises involving major powers during the ninety-year period from 1895 to 1985. A number of historical sources were consulted (see list at end of References) to generate a new data set of major power deterrence cases, which provided the grounds for the empirical analysis. Those cases of general deterrence failure that escalated into immediate deterrence crises between major powers are also summarized in appendix A.³

This book provides both a theoretical and rigorous empirical analysis of deterrence between major powers. The first three chapters (part I) introduce key debates in deterrence research and clarify the main phenomenon that this study attempts to explain. In this introduction, I present alternative theoretical explanations of deterrence and conflict. In particular, the main arguments of those theories that focus on either relative power or manipulative strategies of costly signals are compared to the approach emphasizing the importance of inherent credibility. Later in the book I expand the notion of inherent credibility tying it to regional stakes. This conceptual expansion provides a criterion for identifying major powers' national interests in extended deterrence. As this study is looking at the behavior of major powers, chapter 2 and appendix B provide a theoretical rationale for their ongoing relevance to international relations research. Conceptual criteria and historical background for identifying major powers during the observed period are also presented in detail. Chapter 3 further explores a number of conceptual and operational problems involved in any deterrence analysis and introduces the new data set of deterrence cases.

Each chapter of part II gives a detailed theoretical discussion and empirical test of the individual impact of key factors on deterrence. These include relative power (chap. 4), regional stakes (chap. 5), and domestically driven stakes (chap. 6). Chapter 7, dealing with manipulative strategies of commitments and costly signals, is included in part III because it also probes the possibility of these factors' interactive effects through a series of combined models. As the empirical results point to the vital role of regional stakes for understanding the dynamics of major power relations, whether individually tested or in combination with other factors, part III concludes with a discussion of their implications for conflict theory and deterrence policy, particularly in the post-Cold War period.

The Puzzles of Deterrence

Most theories of international conflict and deterrence that we explore and implement today were formulated or reformulated in the nuclear context of the Cold War. They were meant to provide a logical roadmap for an age in which world stability and survival appeared to rest upon “the irrationality of the rational” logic of nuclear threats. The archetypal situation was that of two nuclear states facing each other with threats of mutual annihilation if either should launch an attack. If a side’s home territory was attacked, the response would be less uncertain since its survival would be at stake. The precarious nature of nuclear deterrence arose, however, when such threats were expected to deter attacks on a third nation, usually the deterrer’s ally or protégé, that is, in the situation of “extended deterrence.”⁴

In the nuclear setting of mutually assured destruction (MAD), the inherent value of a protégé for the deterrer’s national interests becomes a secondary question as this value could not possibly compensate for the costs the deterrer would have to sustain if its threat were carried out. Consequently, the problem of nuclear threats primarily concerns their credibility, which “depends on being willing to accept the other side’s retaliatory blow. It depends on the harm *he* can do, not on the harm *we* can do” (Kahn, cited in Freedman 1989, 134). To solve the credibility problem, i.e., to make irrational nuclear threats look rational,⁵ attention centered on the techniques of signaling intentions. This new focus, in turn, gradually overshadowed the issue of whether the protégé was vital for the deterrer’s national interests in the first place.

Hence, the deterrer of the nuclear age faced the difficult task of persuading its adversary that it would be willing to endure costs disproportionate to the issues at stake whenever its international commitments were challenged. Otherwise, its reputation for strong resolve would be compromised, and the adversary was expected to interpret the deterrer’s threats as bluffs in other places as well. To prevent such a chain of events, triggered by the failure to stand firm, the deterrer had to effectively signal a strong commitment to retaliate, even at the cost of its own annihilation. Would the United States be willing to trade Washington or New York for Paris or Bonn? This was not just a rhetorical question Charles de Gaulle asked at the height of the Cold War.

A number of modern strategic studies, born out of this puzzle, are still in the process of solving it. The questions being asked, and the carefully analyzed answers, delve into the best strategies for signaling foreign policy interests, especially when those interests are weak. Yet, it

appears to be forgotten that the interest in costly signals and reputation was predicated on a very specific idea: that it was *not rational* for an adversary to believe that the deterrer's interests were worthy of retaliation. While it is clear why this was the quintessential premise of nuclear deterrence analysis, it is questionable whether it is useful as a general theory of international conflict. Nevertheless, this approach finds advocates, both in theory and policy, even for nonnuclear deterrence encounters. Moreover, it seems that a reputation for standing firm itself often becomes the key aspect of national interest. Such an approach can almost completely marginalize the critical question of politics, that is, which *political goals* are at stake in specific situations, apart from the reputational issue.

The dilemma regarding relative power capabilities that provide an effective deterrent is another issue that has dominated the conflict and deterrence research. In strategic studies, the emphasis has been on the precarious nature of deterrence brought up by the advent of nuclear weapons. The idea that the strategic balance between superpowers, forged by their second-strike capabilities, assured deterrence stability, found both its advocates and skeptics. Critics of the MAD strategy pointed to its inadequacy to respond to limited attacks or solve the credibility problem. Instead of relying on the strategic balance of terror, they advocated nuclear superiority as a means of solving the credibility problem of nuclear threats in a nonsuicidal manner. The disagreement over the adequate balance of nuclear power that can provide an effective deterrent roughly corresponds to one of the major debates about the relationship between power distribution and war in the general international relations literature.

The realists are divided over the question of whether power balance or imbalance is more likely to lead to deterrence failure and ultimately to war. However, whether considered in the nuclear or nonnuclear setting, the analysis of relative power alone can provide only partial understanding of conflict behavior. It misses the motivational aspect or, in other words, the important factor of each side's willingness to fight over an issue. An asymmetry of motivation, for instance, can reverse the impact of capabilities on the probability of war. Theories and strategies that rest on simple power calculations, without considering motivation or willingness to fight, are at best only incomplete.

It is not by any means argued here that we should forsake the analysis of relative power or its undeniably significant impact on world politics. This study is also not suggesting a minor role of communication and bargaining strategies in crises. In fact, the use of adequate bar-

gaining strategies has been shown to de-escalate even the most acute international crises. Rather, the issue that is raised here points to the danger of applying the logic of nuclear deterrence to nonnuclear situations. The “irrationality of the rational” premise behind nuclear threats is largely unsuitable for nonnuclear milieus. Evidently, deception and costly signaling were considered by some strategic thinkers to be workable solutions to the problem of “unthinkable” scenarios implicit in nuclear threats. However, their unquestioned generalization to other conflict situations, ones void of the irrational logic of nuclear threats, is potentially erroneous and even dangerously risky.⁶ Moreover, the focus on signaling techniques or simple power calculations—at the expense of understanding the inherent issues at stake—obscures the fundamental question of what constitutes vital national interests. Even in cases of nuclear deterrence, the question of national interests needs to be clarified before commitments are made and subsequently honored. As previously discussed, it is reasonable to assume that there is a connection between a deterrer’s national interests and an opponent’s perception of its threat resolve.

The Requirements of Deterrence and Conflict Avoidance

Traditionally, strategic thinkers were engaged in the study of deterrence, while international relations scholars focused on the causes of conflicts and their avoidance. It is reasonable to argue that separation of subject matter, however, is simply artificial, disguised by differences in terminology. The question of what constitutes the “requirements of deterrence” (Kaufmann 1956) is in fact similar to the problem of conflict avoidance. The factors that presumably contribute to deterrence failure can also be considered as the same factors leading to crisis escalation and, ultimately, the outbreak of war. Accordingly, an analysis of the requirements of deterrence should be seen as a special case of conflict escalation theory if deterrence fails and, at the same time, as a theory of conflict avoidance if deterrence succeeds.

The traditional view of the necessary conditions for successful deterrence is that the potential attacker must perceive the deterrer’s threats of retaliation as sufficiently capable and credible for retaliation. Simply put, the potential attacker needs to be *convinced* that the deterrer *can* and *will* execute its threats if the attack occurs.⁷ “Deterrence requires a combination of power, the will to use it, and the assessment of these by the potential aggressor. Moreover, deterrence is a product

of those factors and not a sum. If any one of them is zero, deterrence fails" (Kissinger 1957, 12).

There are, therefore, three basic conditions that need to be met to avoid deterrence failure and crisis escalation. The first requirement is related to the deterrer's retaliatory *capability* and its effectiveness as a deterrent. Regarding power factor, two questions have dominated the deterrence literature: (1) Can an attack be deterred under power parity conditions? and (2) Do nuclear weapons provide an effective deterrent? The second requirement is related to the issue of a deterrer's willingness to carry out its threat. Combined with the third requirement, that the potential aggressor believes the deterrer is indeed willing to retaliate, this defines the notion of *threat credibility*.

Power Capability

Deterrence strategists were principally interested in the limits and potentials of nuclear weapons as deterrents. In this respect, two schools of thought have emerged, one believing the strategy of mutual assured destruction (MAD) to be a sufficient deterrent, the other supporting the development of tactical nuclear weapons to overcome the credibility problem of the MAD strategy. The former theorists believe that the horror of mutual catastrophe works as a strong disincentive for escalatory behavior by either side. Since it makes the use of nuclear weapons unthinkable, the maintenance of a second-strike capability by both nuclear powers is seen as a sufficient deterrent. Or, as Waltz put it: "Nuclear forces are useful, and their usefulness is reinforced by the extent to which their use is forestalled" (1979, 187). On the other hand, tactical nuclear weapons, intended to be used for the purpose of *winning* a nuclear war, are considered dangerous and too risky. They are seen as dangerous because they might tempt adversaries to eventually use them and make "the unthinkable thinkable," to use another of Kahn's phrases.

Contrary to this theory, Kahn and a few others believe in the alternative strategic doctrine, which is expected "to solve the 'credibility' problem of the US deterrent by making it possible to fight a nuclear war in a non-suicidal manner, rather than by manipulating perceptions" (Freedman 1989, 372). In this view, the credibility problem associated with nuclear weapons can be overcome precisely through the development of tactical nuclear weapons, which would be used for limited retaliatory strikes (e.g., Kahn 1965; Gray 1979).

Finally, there is a third approach, according to which nuclear weapons do not have any uniqueness in their deterrent effect and thus are not substantially distinguishable as deterrents from conventional weapons. This argument does not appeal to intuitive thinking about the nuclear age, but there is, nonetheless, a strong record of empirical evidence supporting it. In a number of empirical studies, the possession of nuclear weapons by one or both sides was not strongly correlated with deterrence success or failure (Organski and Kugler 1980; Kugler 1984; Huth and Russett 1984; Geller 1990; Huth 1990). Although a few studies found nuclear weapons to have an effect on the dynamics of deterrence (Bueno de Mesquita and Riker 1982; Huth, Gelpi, and Bennett 1993), it is striking that the deterrent role of nuclear weapons is not as strong as assumed by generations of strategic thinkers. These empirical findings are reinforced by the results from a recent formal-theoretic analysis of deterrence that finds the costs of war beyond a certain point have no impact on deterrence stability (Zagare and Kilgour 2000, 292, 304). Simply put, nuclear states do not behave any differently than nonnuclear states.

Since the empirical evidence does not strongly validate the uniqueness of nuclear deterrence over other types of deterrence, it is plausible to consider the capability requirement for successful deterrence in connection to classical arguments about favorable distributions of power. That is, theoretical explanations of conventional wars as the outcomes of destabilizing power distributions can also be generalized to deterrent relations in the nuclear age. It is not then unusual for analysts to conclude that “the concept of mutual deterrence is, in sense, the classical notion of balance of power in modern disguise” (Dougherty and Pfaltzgraff 1997, 376).

It follows that the logic of the nuclear debate largely corresponds to the essential nature of realist disagreements over what constitutes a more stable power distribution. The “balance of power” approach contends that power parity provides stability as each side faces the same odds of winning or losing a war. An alternative argument, advanced in the “power shift” literature, considers uncertainty about the outcomes of war, created by power balance, as a critical situation that can lead to the outbreak of war. According to power shift school, weaker states are unlikely to engage in conflicts against more powerful opponents, their chances of prevailing over opponents becoming feasible only when they reach a power balance. This, in turn, makes them more willing to initiate conflict or resolutely respond to the opponent’s

challenge. Chapter 4 provides a thorough presentation of each school of thought and then proceeds with an empirical test of their validity for explaining the dynamics of major power deterrence and conflict.

Credibility and Resolve

While major powers can achieve, and often do have, the capability necessary for effective threats, their core problem consists of conveying their willingness to execute their threats. As Kaufmann (1956) and many others have indicated, the crucial requirement for a policy of deterrence lies in the realm of intentions. In addition, an adversary needs to be convinced that these intentions are credible. Hence, deterrence is described as a psychological relationship in which the Challenger's perceptions of the deterrer's willingness to execute threats matter most for deterrence stability or instability.

If the core problem is identified strictly in terms of the perceived credibility of threats, it is not surprising that the deterrence literature has been greatly influenced by a theory of commitments, concerned with the manipulation of an adversary's perceptions of the deterrer's resolve and commitment to act (Schelling 1960, 1966). Still, the view that a threat's credibility is chiefly subjective can fail to recognize some essential elements of deterrence. If the deterrer does not find it worthwhile to fight a war, but attempts to convince the opponent that it does, why would the opponent believe that the deterrer's threat is credible? Similarly, if the deterrer feels strongly about the issue at stake and is ready to fight over it, why would it assume that the opponent would be persistently ignorant of such a strongly held intent? Why not simply communicate what is an already credible intention (if it is truly credible in the first place)?

The approach stressing the manipulative aspects of deterrence strategy, dominated by the theory of commitments as its offspring, leaves these questions unanswered. Even forty years ago Kaufmann (1956, 20) recognized that "there must be some relationship between the value of the objective sought and the costs involved in its attainment. A policy of deterrence which does not fulfill this requirement is likely to result only in deterring the deterrer." For these reasons, as Stephen Maxwell suggested, it is more appropriate to substitute *interests* for *commitments* in the theory of deterrence. The rationale is simple:

The obvious conclusion to be drawn from the failure of a state to fulfill a commitment is simply that the commitment did not repre-

sent an interest worth defending, at the level of violence and risk estimated to be necessary. Thus, if the Americans backed down over Berlin . . . [t]he only conclusion the Russians could confidently draw was that Berlin was not a vital national interest to the Americans. (1968, 19)

Published at the height of the Cold War, when strategists turned to refining the crafts of deterrence strategies, Maxwell raised an issue that has not been addressed systematically, even with the end of the Cold War. Maxwell was not alone, though. According to Jervis, the “third-wave” theorists of the 1970s began to recognize the importance of “intrinsic interests,” that “represent the *inherent* value that the actor places on the object or issue at stake” (1979, 314; emphasis added). Basically, we can identify two different schools of thought concerning threat credibility: commitment theory and theory of inherent credibility.

Although the fundamental differences between these theories are rarely registered in the literature, their logic roughly corresponds to Huth’s (1997) distinction between “strong-interdependence-of-commitments” and “case-specific” credibility.⁸ Commitment theory can be traced to the pioneering work of Schelling, Kahn, and others. Since it is the single most influential theory of deterrence, it should be no surprise to learn that it has received much more attention in the literature than have theories of inherent credibility. Summaries of this classical approach to deterrence can be found in Young (1967), where it is labeled “manipulative bargaining theory,” and Zagare (1996) and Zagare and Kilgour (2000), where it is called “decision-theoretic deterrence theory.” Note that here the term *commitment* does not refer strictly to formal alliances, but it rather encompasses an entire range of techniques for signaling strong resolve regardless of the involved inherent stakes.

Commitment Theory

In their seminal studies *The Strategy of Conflict* (1960) and *On Escalation* (1965), Schelling and Kahn, respectively, paved the way for the modern strategic studies that have contoured much of our thinking about deterrence and conflict in the nuclear age. Kahn thought that the development of tactical nuclear weapons, for possible limited use, was one way to solve the nuclear deterrence problem. At the same time, he was interested in the question of achieving deterrence through strategies of firm commitments and costly signals. Still, it is unquestionable that commitment theory is mostly the product of Schelling’s influential

work in this area (1960, 1966). Schelling essentially argued that the problem of nuclear threat credibility should be solved through the refinements of signaling techniques that manipulate the levels of risk. There are a number of ideas that follow from Schelling's work on "the art of manipulation."

One of the major rationales for maintaining a reputation for strong resolve comes from the understanding that commitments and events are seen as interdependent (Schelling 1966, 55–59). In this view, the interdependence of events takes two forms. First, commitments and events are considered to be horizontally (spatially) coupled, implying that a state's failure to react to challenges in one region would diminish other states' perception of its resolve to honor commitments in other regions as well. The domino theory that guided Cold War containment policy was clearly premised on the understanding of international events as spatially interdependent. The other form of interdependence is related to the past behavior and the consistency of strong resolve over time, indicating what we may call the vertical (temporal) interdependence of commitments. Its metaphorical version, the "Munich analogy," figures prominently as a driving force behind a number of foreign policy decisions. Scholars in this tradition argue for the validity of both approaches to the interdependency of events, succinctly summarized by Schelling: "Few parts of the world are intrinsically worth the risk of serious war by themselves . . . but defending them or running risks to protect them may preserve one's commitments to action in other parts of the world at later times" (1966, 124).

As for the nature of threats, Schelling's approach develops an entire range of techniques for signaling strong commitments. The key idea is that "it does not always help to be, or to be believed to be, fully rational, coolheaded, and in control of oneself or one's country" (Schelling 1966, 37). Essentially, it is consequently assumed that the deterrer benefits by leaving the impression that the process is not completely controllable (Schelling 1960, 187–203; 1966, 35–91). "The key to these threats is that, though one may or may not carry them out if the threatened party fails to comply, *the final decision is not altogether under the threatener's control*" (Schelling 1960, 188; emphasis in the original). The idea is comparable to the "rationality of the irrational" strategy (Kahn 1965) that consists of compulsive or unreliable behavior, such as teaming up with a compulsive ally or delegating the authority to use nuclear weapons to lower-level commands (Schelling 1966, 36–43).

Besides recklessness and unpredictability, such as compulsive or

unreliable behavior and the “threat that leaves something to chance” (Schelling 1966, 36–43; 1960, 187–203), the deterrer’s threat can be greatly enhanced by costly signals. This notion is closely related to Schelling’s understanding of deterrence as “competition in risk-taking.”

The creation of risk—usually of shared risk—is the technique of compellence that probably best deserves the name of ‘brinkmanship.’ It is a competition in risk-taking. It involves setting afoot an activity that may get out of hand, initiating a process that carries some risk of unintended disaster. (1966, 91; see also 1960, 199–201)

“Bridge burning” is yet another idea in the catalog of costly signaling techniques. “Often we must maneuver into a position where we no longer have much choice left. This is the old business of burning bridges” (Schelling 1966, 43). Routinely attributed to Schelling, the notion of burning bridges is actually equivalent to Ellsberg’s earlier idea that the chances of successful deterrence are enhanced if a deterrer (“blackmailer”) “voluntarily but irreversibly gives up his freedom of choice” (1959, 357). Again, the essential message to be conveyed to an adversary is that the process of retaliation is not fully controllable.

Deterrence theorists who subscribe to the signaling approach generally examine and refine a number of ideas generated from Schelling’s works on “the art of manipulation” (e.g., Jervis 1970; Powell 1990; Fearon 1997). Fearon’s formal analysis, for instance, adds refinements by distinguishing between the “tie-hands” and “sink-cost” strategies. He finds that signaling states do better by tying their hands, that is, by creating costs that would be paid *ex post* if they fail to honor their commitments, rather than by using a sink-cost strategy (Fearon 1997, 82). The tie-hands strategy is linked to the issue of *ex post* costs arising from the actions of domestic political audiences. Subsequent formal stylizations have added more insights into the credibility problem as a function of domestic audience costs (e.g., Fearon 1994b; Schultz 1998; Smith 1998). On the other hand, the sink-cost strategy is costly *ex ante* for the deterrer. “Leaders might . . . sink costs by taking actions such as mobilizing troops or stationing large number of them abroad that are financially costly *ex ante*” (Fearon 1997, 82). It is the sink-cost strategy, therefore, that fits more closely the classical commitment theory as defined here. As already discussed, its central idea is that events are interdependent rather than discrete. For this reason, costly threats such as troop mobilization or even limited interventions are deemed

necessary in order to project a reputation for strong resolve and prevent domino effects elsewhere. This is exactly what Fearon (1997) identified as a sink-cost strategy as distinct from the one that ties hands through the ex post domestic audience costs.

To what degree is commitment theory empirically valid? It has certainly been influential in politics, a number of strategic thinkers observing that, during the Cold War, the United States was indeed greatly concerned with its reputation for fulfilling its commitments. According to several historical analyses, this concern was found to be so strong that it was often given primacy regardless of the specific contextual issue involved in the crisis (George and Smoke 1974; Betts 1987). It is interesting to note that the same studies did not find other major players following a similar logic (e.g., Betts 1987, 13; Mercer 1996), not even the Soviet Union (Jervis 1985; Hopf 1994), or an ally such as France (Weinstein 1969). The approach of other major powers to deterrence appeared to be rather context-dependent, fitting more the idea of what was inherently at stake in a particular situation. The goal of honoring commitments, irrespective of the particular context, seemed to play only a secondary role in their behavior.

Besides the question of the degree to which states follow costly signaling prescriptions, there is also the empirical question of whether such strategies work as deterrents. The empirical findings are not quite positive in this respect. In their comparative case studies, George and Smoke (1974) did not find that communicating signals visibly affected potential attackers.⁹ Instead of the strategy of “burning bridges” and “leaving something to chance,” George and Smoke recommend a strategy that gives an adversary multiple options for response. Quantitative empirical studies also show that a successful resolution of international crises requires, among other things, less ambiguity achieved through overt communication of intentions (Leng 1993; Dixon 1996). In other words, bargaining strategies do matter, but not necessarily through the strategies of risk manipulation, costly commitments and signals, or foreclosing the options for response.¹⁰ As for the effect of past behavior on the chances for successful deterrence in current or future conflicts, Huth reports mixed findings. “Weakness in one situation did undermine a state’s future credibility; . . . But firmness in one situation did not necessarily enhance a state’s credibility in another” (1988, 81).

Nevertheless, the concern with international reputational costs appears to have guided U.S. policy in a number of Cold War and post-Cold War crises alike (e.g., Layne and Schwarz 1993; Layne 1997). As an illustration, U.S. credibility was listed at the top of the

administration's reasons for the invasion of Haiti: "First is the essential reliability of the United States and the international community. . . . we must make it clear that we mean what we say. Our actions in Haiti will send a message far beyond our region—to all who seriously threaten our interests."¹¹ Despite the post-Cold War environment in which the Haitian intervention took place, such a rationale for the intervention is hardly distinguishable from the logic of domino theory that guided U.S. Cold War policy: "if we were to let Asia go, the Near East would collapse [and] no telling what would happen in Europe" (Truman, cited in Jervis 1991, 20).¹² The obvious impact of reputational concerns on U.S. strategic thought and policy after the Cold War has not faded away even though empirical findings point to the need for substantial modifications in this kind of thinking. Based on these findings, it is plausible to suggest that the question of whether vital interests are at stake needs to be carefully examined first, before any costly signals are sent to an adversary. Such policy advice is more consistent with an alternative school of thought, one centered on the notion of inherent credibility rather than the manipulative strategies of commitments and signaling resolve.

A Theory of Inherent Credibility

In the 1960s and 1970s, several third-wave deterrence theorists (see Jervis 1979) raised doubts about the validity of the theory of costly signaling and commitments. Its suitability was questioned for both nuclear and conventional threats. Maxwell (1968) was one of the first to observe that it was difficult to understand why it would be rational for a state to engage itself in a nuclear war over an issue it did not consider worth a war "simply in order to support the credibility of its commitment to other objectives it did consider worth waging a war. To do so would mean deliberately exchanging a *risk* of nuclear war, even if a very high one, for a *certainty* of nuclear war." The most important inference drawn from his criticism is that "if interest is substituted for 'commitment,' the argument for interdependence immediately becomes less plausible" (12, 19).

Even Schelling, as a first-wave theorist, was aware of the importance of the inherent stakes or interests that the opponents had in a deterrence situation. He originally made a distinction between "warnings," intended to convey a deterrer's "true incentive structure," and "threats," conveying the intention of changing it (1960, 123–24 n. 5). Because he focused on threats and not warnings, later works mostly interpreted the credibility of deterrence in terms of the deterrer's effec-

tiveness in projecting its resolve to change its own “true” or “intrinsic” values placed on the protégé, particularly when they were inherently weak. This is exactly what Kaufmann called “detering the deterrer” rather than deterring the potential attacker.

Third-wave theorists, in fact, restated this distinction, but shifted their attention to the sources of Schelling’s warnings. Following Maxwell’s forceful essay about the hazardous aspects of commitment theory, Weinstein (1969) distinguished between “situational” and “nonsituational” concepts of commitments. The former was inherent to the situation, regardless of verbal or nonverbal signals. Weinstein argued that it was precisely the implicit “situational” commitments that were better predictors of behavior than verbal (“nonsituational”) commitments, since they were dictated by the national interests involved in the crisis (43).

Jervis (1970) also made a distinction between “signals” and “indices,” respectively analogous to Schelling’s “threats” and “warnings.” “*Signals* are statements or actions the meanings of which are established by tacit or explicit understandings among the actors. . . . signals are issued mainly to influence the receiver’s image of the sender . . . [but] they do not contain inherent credibility.” In contrast to signals, indices “carry some inherent evidence that the image projected is correct” (18). Jervis defines indices as a comprehensive category, which may include a leader’s personal behavior, a state’s domestic political system, or power capabilities. They can also be verbal statements, unless the receiver believes they are meant to influence him, in which case they are classified as signals (30–39). Snyder and Diesing (1977) followed with their distinction between “inherent bargaining power,” on the one hand, and “actual bargaining power” and “bargaining skill,” on the other. The latter were a function of a challenger’s perceptions and the deterrer’s manipulation of its own “inherent resolve,” whereas the former represents the “inherent” resolve that “sets a certain limit on what the parties can achieve by manipulative tactics” (Snyder and Diesing 1977, 190).

It appears that Jervis’s work on deterrence eludes easy classification into one or another theory,¹³ though he did make an argument that the “indices” made the greatest impact on potential initiators (1970, 18, 65). This argument was stated even more strongly by George and Smoke, after their careful examination of a number of deterrence cases in American foreign policy. “The fact of the matter is that the task of achieving credibility is secondary to *and dependent upon* the more fundamental question regarding the nature and valuation of interests” (1974, 559; emphasis in the original).

Most deterrence theorists did not move beyond these basic conceptual distinctions. As a result, they continued to view the credibility of deterrence in terms of effective manipulation (Schelling's threats, Jervis's signals, Weinstein's nonsituational commitments, and Snyder and Diesing's bargaining skills). In a series of quantitative deterrence studies, Huth and Russett (1984) revived interest in the issues raised by third-wave theorists. Their results provided additional justification for a better understanding of the "issues at stake," that is, the less manipulable inherent interest in the issue of dispute. Their findings were not uniform for all indicators of interests at stake, but were nonetheless critical for the advancement of deterrence theory. In general, they provided empirical validation for the importance of inherent interests, apart from manipulative strategies, for the calculus of deterrence.

Since inherent credibility can set limits on the effectiveness of risk manipulation, it is arguably an even more salient type of credibility. In this study, I provide an extension of the idea of inherent credibility, identifying it with the traditional notion of vital national interests. Previous work on extended deterrence focused exclusively on the isolated value of a Protégé to a potential Defender. However, even if a particular state does not have great individual significance for a major power, it may still be important if it is located in a region of critical strategic importance to that major power's interests. For this reason, I suggest that a Defender's vital national interests may be better understood by looking at the regional level of major power rivalry. The approach developed here provides a measurable indicator of national interests for major powers—apart from those directly affecting the homeland territory—by identifying them with the variable stakes they might have in different world regions.

It is the search for better understanding of external sources of credibility that sets this analysis apart from the voluminous recent literature on domestic sources of foreign policy resolve. Domestic politics has been examined principally in terms of the impact of the regime type (assumed to be correlated with the domestic costs for foreign policy failures) on a deterrer's resolve. My model of external (regional) sources of threat credibility is not incompatible with the domestic politics models. Nevertheless, we may compare their relative explanatory power as will be done in chapters 6 and 7.

Theoretical Explanations

A central purpose of this work is, therefore, to explore the neglected issue of the effects of inherent credibility on deterrence stability. My

main argument is that vital national interests are intrinsically linked to the idea of inherent credibility, and that the question of what constitutes these interests should form the core of any examination of deterrence credibility. This question has not been sufficiently explored in the literature, nor is it always clearly understood in the actual conduct of foreign policy. The theoretical approach guiding my analysis does not view inherent credibility as an isolated factor. Rather, it points to the importance of the interplay between inherent credibility, resolve, and capabilities for explaining the dynamics of deterrence. The argument is analyzed in regard to major power relations, but it is logically generic and applicable to the behavior of other states as well.

As for the importance of major powers, it has been empirically demonstrated that they have been the most frequent participants in international crises and wars (Bremer 1980, 79; Maoz 1982, 55; Small and Singer 1982, 180; Gochman and Maoz 1984, 596; Huth 1998, 744–45). The empirical evidence also shows that their involvement tends to both increase the probability of crises escalating into wars and heighten the severity of wars. Threats to a major power's territory are rare, and most major powers react militarily and resolutely to any direct attack on their territory, as demonstrated by the U.S. response to the Japanese attack on Pearl Harbor (1941) and the Soviet reaction to the German blitzkrieg (1941).

Major powers are therefore less likely to directly threaten each other's territory than to clash over their interests regarding third nations, such as allies, major economic partners, or other protégés in regions vital to their interests. Examples of major wars resulting from extended deterrence failures are numerous: the Russo-Japanese War (1905) escalated from a dispute over Manchuria and Korea, World War I (1914) escalated from the Balkans crisis, and the German invasion of Poland triggered World War II (1939). Such disputes over third parties are common, yet we do not fully understand why only some of them escalate into major confrontations.

To be precise, questions that still require adequate answers in the deterrence literature can be identified as follows:

- (1) When will a major power (Challenger) decide to become involved in a conflict with another nation (Protégé)?
- (2) When will another major power (Defender) react to defend the Protégé against the Challenger?
- (3) Faced with the demand to withdraw from its conflict with the Protégé, how likely is it that the Challenger will back down?

- (4) If the Challenger persists in its conflict with the Protégé, will the Defender carry out its threat?

My analysis attempts to answer these key questions of deterrence by estimating the explanatory power of four factors. Each factor is first examined individually, then their potential interactive effects on deterrence are explored in a series of combined models (chap. 7). These factors, providing four alternative explanations of deterrence, are as follows:

- (1) the capabilities for carrying out a threat, defined in terms of power distribution between major powers at the moment of their dispute over a third nation;
- (2) the inherent credibility of each side's threats, reflecting the stakes that the confronting major powers have in the region of the third nation's location;
- (3) the resolve of each side to carry out their threats, assumed to be a function of the domestic costs they are likely to endure as a result of foreign policy failures;
- (4) the international reputation for honoring commitments developed through the pursuit of costly signaling strategies.

The book ties the analysis of the two basic requirements for successful deterrence, threat credibility and retaliatory capability, to the analysis of the causes of escalation and war. Retaliatory capability is linked to the analysis of relative power distribution, while threat credibility in this framework is a function of three alternative factors: (a) the intensity of linkages between a major power and the region in which a conflict occurs (an extension of inherent credibility theory); (b) the domestic incentives for not backing down; and (c) the pursuit of a costly signaling strategy intended to develop an international reputation for strong resolve in honoring commitments (i.e., the core of commitment theory).

Power Distribution

Of all factors, distribution of power has been the most widely analyzed factor in the literature. It gave rise to a major debate between those who believe that a balance of power constrains opponents from attacking each other and those who argue that power parity gives states incentives for using force, as either side might believe it has the chance

to win a war. I address this debate by exploring the question of which power distribution leads to more violent or peaceful resolutions. In deterrence, it translates into the question of whether deterrence is likely to fail under a balance or imbalance of power between the major powers involved. Chapter 4 presents and tests the validity of such theories, whereas chapter 7 compares their validity to the explanatory power of the other three factors and also assesses the possibility of their interactive effect on deterrence. In light of the previous discussion and empirical findings, the presence of nuclear capabilities is assumed to be relevant, but not critical for deterrent effects. The framework is intended to explain the logic of deterrence in general, whether it be nuclear or conventional. As such, it provides the basis for generalizations beyond Cold War dilemmas.

Regional Salience or Stakes

As already discussed, threat credibility has routinely been examined in terms of choosing the best manipulative strategies for convincing an opponent that a threat will be carried out, even when one's commitment to defend the third party is very weak. I take a different approach by emphasizing the importance of so-called inherent credibility or intrinsic interests, thus picking up where third-wave deterrence theorists left off in their conceptual discussions (i.e., George and Smoke 1974; Snyder and Diesing 1977; Jervis 1979). My research also expands, both conceptually and theoretically, the notion of the issues at stake explored empirically in quantitative deterrence research (Huth 1988; Huth and Russett 1984, 1988, 1990).

To be precise, *inherent credibility* refers to the less manipulable interests or stakes that forge a major power's willingness to carry out its threat. The question of how nations signal their weak and questionable stakes as credible commitments to defend a third party is considered irrelevant. Rather, it is essential to understand the conditions under which the commitments are likely to be strong in the first place, hence giving rise to inherently credible threats. If a third nation is located in the region of a Defender's vital interests, I argue that the Defender will be *willing* to carry out its threat against the Challenger. The Challenger has the option of backing down, which I predict will happen if its stakes are not as high as those of the Defender. When the stakes are high for both sides, the probability of war between major contenders is heightened.

In chapter 5, I introduce and examine a new observable category of inherent credibility, consisting of regional interests of major powers in the area of dispute. Specifically, I provide a linkage between the notion of inherent credibility and the geopolitical literature on regional aspects of major power rivalry. My general assumption is that the more salient an area is for a major power, the more likely that power is to become involved in a crisis arising in the area. If an area is highly salient for more than one major power (i.e., it is an area of conflicting interests, also known as a gray area), the likelihood of a more serious conflict between major powers increases. This view is consistent with the intuitive claims made by a number of diplomatic historians and geopolitical writers, but hardly ever rigorously tested in previous research. In order to test my argument, I developed a new measure for the salience of regions to major power interests and data were accordingly gathered for the observed period from 1895 to 1985.

My theoretical argument about the relevance of regional stakes for understanding the credibility problem in major power deterrence is first examined apart from other factors. It is then combined with the capability variable—another key requirement of deterrence—to examine if there is an interplay between power distribution and regional stakes that can explain a good part of the variance in deterrence outcomes.

Domestic Audience Costs

The disputants' domestic regime type can be considered as the second important source of threat credibility. Although a few early deterrence writings considered threat credibility to be at least partially a function of domestic public opinion (e.g., Kaufmann 1956), this idea was not integrated into deterrence analyses until very recently. Recent formal stylizations of the impact of "domestic audience costs" on a state's behavior have revived the issue of domestic politics in deterrence research (e.g., Fearon 1994b; Schultz 1998; Smith 1998). The models examined the domestic costs of deterrence for a Defender, foremost the costs of perceived foreign policy failure, as the critical indicator of a Defender's resolve and of the credibility of its threats. A Defender's vulnerability to its domestic audience demands is believed to be a function of its regime type, democratic leaders being more vulnerable to domestic audience costs than the leaders of nondemocracies (Fearon 1994b, 585). Hence, one of the key arguments is that democratic gov-

ernments' threats have more credibility than those of their nondemocratic counterparts, as democratic governments face higher domestic political costs for not carrying out their threats.

The argument is novel and appealing for its intuitive implication that resolve in deterrence is tied to the domestic milieu, and that costs of action or inaction subsequently arise from the interplay between domestic and foreign policies. Note that such a claim does not imply the notion of inherent credibility, though it also does not require any manipulative behavior for deterrence to work.¹⁴ As such, it does not belong to the research tradition emphasizing the importance of inherent credibility, yet neither does it contradict this tradition. Like the other three variables, the impact of domestic regime type—reflecting different levels of costs for leaders' foreign policy failures—on deterrence outcomes is first examined by itself. Chapter 6 addresses the question of whether the incentive for not backing down in deterrence crises is shaped by the domestic environment. In the following chapter, this variable is incorporated into the combined models to explore the robustness of the findings from chapter 6 and the possibility of interactive effects with other key factors.

Commitment Strategies and Costly Signals

As previously discussed, commitment theory of credibility rests on the assumption that commitments are spatially and temporally interdependent. The failure to stand firm on one issue is assumed to erode a deterrer's reputation for strong resolve on other issues. Disputes can arise in different geographic areas or time periods. Regardless, this approach contends that opponents infer a deterrer's resolve from one event to predict its behavior in another event, despite the spatial or temporal distance between their occurrences.

Chapter 7 provides an empirical test of this view by focusing on two explanatory factors that are central to commitment theory. First, I explore the deterrent effects of the strategy of costly signals in terms of the behavior that creates *ex ante* costs for the deterrer (i.e., a "sink-cost" strategy). Second, I examine the impact of a deterrer's past behavior on the rate of its success or failure in subsequent deterrence crises. As the findings will show, inherent credibility, interpreted to reflect vital interests in the region of dispute, provides better explanatory power than either of these two key elements of commitment theory.

In summary, this book examines the dynamics of deterrence between major powers, especially their behavior in military conflicts in which

two or more major powers are disputing over a third nation. This is the most frequent type of major power conflict, yet it has received little attention from scholars. The theoretical background suggests four factors be examined as predictors of the likely outcome of such conflicts: power distribution, regional stakes, domestic regime type, and commitment strategies of costly signals.

Furthermore, the model assumes that regional stakes, a factor previously ignored in deterrence studies, are a more powerful source of inherent credibility of extended threats than other factors, such as relative capabilities or a deterrer's resolve, shaped by the expectations of its domestic costs or international reputation for standing firm. If we were to predict the choices that major powers are likely to make in their disputes over other nations, their involved stakes should be the most potent factor to keep in mind. This central argument of the book brings back into focus the often neglected issue of when and where vital national interests are at stake. The question of national interest should precede all other concerns and be addressed by policymakers *before* embarking on any risky behavior to defend a reputation for strong resolve and toughness for its own sake.

This approach can thus provide a strategic guide for a major power's grand strategy, such as that of the United States in the post-Cold War context. As a number of U.S. observers and scholars have been critical of U.S. foreign policy for its lack of clarity and prioritization of its post-Cold War national interests, this study provides one possible way to move beyond the residuals of "unrestricted globalism" driven by the domino and other Cold War metaphors.