1

Understanding the Design of Security Commitments

The claim that ambiguity can maintain peace in a crisis challenges our intuition. Ambiguous, weak, and fumbling commitments have been blamed for many wars. Most famous, perhaps, is England's vague commitment to France and Russia during the July 1914 crisis. Sir Edward Grey, British foreign minister, refused both to promise neutrality to Germany and to extend security guarantees to Russia and France beyond the loose alliance framework established by the 1904 Entente Cordiale between Britain and France and the 1907 Anglo-Russian Entente. Many claim that, at a minimum, Grey's strategy failed to prevent an avoidable escalation of conflict and may even have caused Germany and Austria, as well as France and Russia, to act on misperceptions about Britain's intentions (Albertini 1957; Snyder 1984; Trachtenberg 1991). Similarly, scholars have argued that weak third-party military commitments to European powers, especially Czechoslovakia and Poland, failed to deter Germany in 1938 and 1939 (Taylor 1961; Morrow 1993). More recently, critics have charged that mixed signals from the United States in 1990 about how it would respond to Iraqi aggression on its southern border opened the door for Saddam Hussein's invasion of Kuwait. Had the United States declared its intention to respond militarily, the argument goes, war may very well have been avoided (Jervis 1994).

Prevailing wisdom maintains that third-party defenders have the best shot at reducing misperception and war if they extend strong, well-defined public pledges of military support to their allies (Fearon 1997; Huth 1999; Schelling 1960; Zagare and Kilgour 2000). However, governments often do not fully flesh out the details of many commitments, including formal military alliances. Of 259 alliances formed between 1816 and 2000 and designed to deter threats to allies, 74 promised to defend the ally no matter what, 139 conditioned third-party intervention on the initiation of conflict by a non-alliance member, and

46 were "ambiguous" in that signatories did not have automatic contractual obligations to intervene on behalf of fellow alliance members in war.[1]

International security commitments are often incomplete because it is simply not possible to anticipate every potential incident covered by the scope of the promise. However, many are deliberately designed to be ambiguous. Consider, for example, the U.S. commitment to defend several small island groupings just off the coast of China during the Quemoy crisis. In 1954, the Chinese Communist Party Leader and Chairman of the People's Republic of China (PRC), Mao Zedong, declared that Taiwan should be "liberated" and began shelling the island of Quemoy, which was held by Chiang Kaishek and the Chinese Nationalists and was a stepping-stone to a PRC invasion of Taiwan. The Eisenhower administration sought to deter further Communist advances on the offshore islands, and especially Taiwan. The textbook strategy for deterrence maintains that the United States should have committed transparently and irrevocably to defend all Nationalist-held territories. Instead, Eisenhower deliberately created uncertainty about whether the United States would intervene to defend the offshore islands. In a formal treaty signed in 1954 between the United States and the Republic of China (the government led by Chiang), the United States agreed to defend Taiwan, the Pescadores, and "such other territories as may be determined by mutual agreement." Secretary of State John Foster Dulles described the strategy as "deterrence by uncertainty" (Chang and Di 1993, 1511). He commented that the treaty "stakes out unqualifiedly our interests in Formosa [Taiwan] and the Pescadores and does so on a basis which will not enable the Chinese Nationalists to involve us in a war with Communist China" (Garver 1997, 114).

Clearly, there was a belief in Washington that firm commitment to the offshore islands posed unwelcome risks, which did not exist for the Pescadores and Taiwan. Members of the Eisenhower administration worried that an overly firm commitment to the offshore islands might enable Chiang to entrap the United States in a war with the Chinese Communists on mainland China, whereas his ability to take advantage of a firm commitment to the Pescadores and Taiwan in a similar fashion was limited. The hesitation to commit to the offshore islands stemmed from the fact the islands were small, located just a few miles off the coast of China's mainland, and scattered along China's long coastline. Additionally, they served as staging points for ongoing skirmishes. Many hostilities targeting and originating from these territories would, therefore, be especially difficult to monitor, making it difficult to assess blame and punish instigators of conflict. Consequently, not only were they difficult to defend from a PRC invasion, but they also presented an opportunity for Chiang to provoke the Chinese Communists.

This example brings into sharp relief the impact of moral hazard in alliances. Moral hazard results when an actor is enticed to behave aggressively because

[1] The typology of alliances developed in Chapter 2 discusses the details of these alliances.

it is insulated from the risks of its actions. It occurs, in this context, when the protégé country, which is the beneficiary of a third party's security guarantee, is emboldened to belligerent action because the third party has promised to provide military support if the protégé is involved in conflict.[2] Consequently, third-party defenders who are motivated to design extended security commitments with the objective of protecting the protégé while preserving the status quo will also worry about the effect of the security commitment on the protégé's behavior. Thus, the third party will design alliances with the protégé's incentives in mind.

The phenomenon of moral hazard lies at the heart of a central tension in forming extended deterrence commitments. A third-party defender wishing to protect a protégé can best deter an adversary's challenges to the protégé by forming a maximally credible and firm commitment, but such commitments risk emboldening protégés not only to resist the adversary's challenges but also to provoke the adversary in an attempt to gain concessions (e.g., more territory). How can leaders deter enemies while restraining allies? This dilemma is a long-standing fascination of international politics scholarship. Glenn Snyder and Paul Diesing (1977, 432) identify the problem as the "deterrence-versus-restraint" dilemma. Robert Jervis calls it "dual deterrence" (1994, 122–124). James Fearon, who states that this is a common historical problem, refers to it as "the problem of moral hazard in alliances and extended deterrence" (1997, 84).

However ubiquitous this conundrum is, the puzzle of how leaders design commitment mechanisms to resolve it remains largely unanswered. This mechanism design problem in extended deterrence and alliance formation can be examined through three fundamental questions. First, how do third-party defenders balance the demands of deterrence with the risks of moral hazard? Second, what does an ambiguous commitment mechanism look like in theory and practice? Third, why do defenders ever opt for ambiguity when transparent alternatives exist? This book offers answers to these questions. The primary thesis is that leaders form many different kinds of security commitments in practice, with much of this variation explained by the moral hazard they face. Furthermore, under certain conditions, a probabilistic commitment is often deliberately incorporated into alliance contracts to deter threats while simultaneously restraining allies.

The Content of Commitments

In addressing the first question regarding how to balance the demands of deterrence with the risk of moral hazard, one of the book's central claims is that the

[2] More generally, moral hazard is a concern in a large and diverse set of contracting environments ranging from decisions about medical care and insurance (Arrow 1963; Shavell 1979), corporate investment, driving behavior (Holmstrom 1979), electoral politics (Banks and Sundaram 1993), and the behavior of athletes (Goff 1997).

content of the commitments in alliance contracts matters a great deal for structuring leaders' incentives. The argument begins with the empirical observation that leaders use a broader menu of contracts to form interstate alliances than is traditionally recognized. Throughout recorded history, leaders have promised other states military assistance, usually expressed in policy pronouncements or formal alliance treaties. Scholars have conventionally categorized these commitments as either "offensive" or "defensive," according to whether the objective is to gain concessions for alliance members from nonmembers or to protect alliance members against nonmembers. However, these categories are not exhaustive, nor do they fully capture the dimensions of variation among interstate alliances.

To illustrate this variation, consider the traditional category of "defensive" alliances, which are formed with the objective of deterring a threat to a protégé. To achieve this goal, an ally may choose to make an unequivocal and unconditional commitment to come automatically to the aid of alliance members, even in the face of moral hazard. Such commitments may involve a pledge of unlimited support or may specify clearly the amount of support to be transferred to the ally. The Soviet Union's alliance with Romania in 1948 is an example of an unconditional commitment to deter German aggression, because the objective of the agreement was to defend alliance members and it did not restrict military intervention to a specific action or condition. At other times, allies write ambiguous alliance contracts, such as Eisenhower's 1954 alliance with the Chinese Nationalists. These contracts leave open the question as to whether the defender will actually intervene if there is war. Another option is a conditional commitment. Leaders often promise military assistance on the condition that the adversary initiates the hostilities. In other words, the third party will often try to limit its commitment to conflicts that do not result from aggression by the protégé. For example, in 1912 the United Kingdom formed a conditional deterrent commitment to protect Belgian neutrality. In the treaty, the United Kingdom stipulated that it would wait until either France or Germany attacked the other before it would intervene to protect Belgium, which sits between the two powers. Conditional commitments occur frequently in alliances and might, under some conditions, help reduce the risk-taking behavior that can result from moral hazard.

To make sense of this rich variation in commitments, the book presents a novel typology of military alliances that facilitates the development of a theory about the distinctions observed in practice. The promises and threats contained in military commitments typically conform to a basic structure with an antecedent condition that, if realized, invokes a consequent obligation. The historical record shows that the terms of both antecedents and consequents exhibit systematic variation, which makes it possible to classify commitments into coherent categories based on these differences. As might be expected, some of the categories of military alliance generated by the typology overlap existing conceptions of commitment, but the typology also uncovers types of

commitments that are not present in theory and yet are routinely adopted in practice.

Many of the alliances considered here, including most of the examples cited previously, have traditionally been thought of as simply "defensive" pacts. However, these alliances, which share the objective of deterring threats to allies, have important categorical differences. Given these differences, a valid question is why previous research has not studied in depth the variation in the content of interstate military alliances, including the variation among extended deterrence commitments. In the existing literature, extended deterrence commitments are usually viewed as having a binary quality – one is committed or one is not – rather than as a mechanism that can take different forms and serve different purposes depending on the conditions and obligations built into the pledge.

The binary viewpoint results from the emphasis on credibility. As in Fearon's (1997) hands-tying and sunk-costs model, commitment is thought to occur once an actor has expended sufficiently high costs to convince others of the credibility of its intentions. This intuition has dominated our thinking about commitment since Schelling (2006) emphasized the importance of irrevocably binding oneself to a future action to attain credibility. Furthermore, scholars have often assumed that states have a strong *preference* to be fully committed to an act. As the examples that open this chapter demonstrate, many scholars hold the view that the lack of a credible commitment has contributed to war. As a result, scholars, pundits, and policy makers alike have theorized how states might make their commitments credible.

The problem is particularly acute in the case of third-party commitments, since it is not immediately credible that a third party will actually provide assistance to another actor when so doing imposes costs on itself. Few people doubt a leader's resolve to defend his or her own country against attack. It is considerably harder to believe that a country's leader and people will deliver on a promise to fight for another country when remaining uninvolved does not put their own safety at risk. Consequently, the dominant research question has been: How can third-party actors bind themselves to the security of another state and foreswear abandoning that state when harm befalls it? With the focus on becoming committed and establishing credibility, it is unsurprising that the dominant concept of commitment is binary, yes or no. This viewpoint makes it difficult to imagine an option other than automatic intervention or nonintervention, such as a deliberately ambiguous commitment.

In practice, however, ambiguous commitments are common. A classic example is the long-standing U.S. policy of strategic ambiguity toward the Taiwan Strait. The U.S. government is unwilling to "become committed" in the conventional sense of establishing a credible threat to defend Taiwan if China attacks. However, U.S. policy is definitely not equivalent to noncommitment. Rather, there is a commitment, which is acknowledged by the United States, Taiwan, and China, but it is not clear whether that commitment obligates the

United States to defend Taiwan automatically if there is war between China and Taiwan.

One possible explanation for this situation is that the Taiwan Strait environment consists of unique circumstances that led to the emergence and persistence of an ambiguous policy in the early 1950s, and the resilient peace between China and Taiwan is unrelated to the U.S. commitment. However, there is another possibility, which is that third-party states routinely and purposely form ambiguous commitments because in many contexts, an irrevocable stand will cause the protected state to react too much. In these instances, ambiguity is a virtue and it is a mistake to view wishy-washiness as a missed opportunity. This is the argument pursued in this book. In the subsequent chapters, I develop general explanations for ambiguous commitments, and I illustrate that under certain conditions, the uncertainty generated by ambiguity has a positive, though counter-intuitive, effect on interstate peace.

Although credibility does play a role in shaping actors' behavior when forming military commitments, this book takes a broader look at the determinants of treaty design. A central contention is that leaders' goals of credibility are tempered, and occasionally overshadowed, by worries about the dangers of overcommitment, which may tempt the protégé state to engage in risky, aggressive behavior. Fortunately for third-party leaders, ambiguous commitment to defend an ally can discourage aggression (or at least too much aggression) because the protégé state now faces doubt about whether its alliance partner will, figuratively, have its back. If ambiguity is sufficient to deter aggression by both the protégé and the opponent, then guaranteed commitment may be inefficient and may even cause behavior inconsistent with the third party's deterrence goals. To explore these matters further, this book deals with the *content* of commitments to support another actor militarily, with a particular focus on understanding why leaders often deviate from the "standard" of full commitment and opt instead for ambiguous or probabilistic commitments that build in discretion for the third party to assist or to abandon according to its preference.

Characterizing Ambiguity as Probabilistic Commitment

The second objective of the book is to flesh out our understanding of what an "ambiguous commitment" looks like in both theory and practice. How and why do states create uncertainty about their commitments to their protégés? The previous section established that states use a variety of types of commitments, including ones that involve ambiguity, in their alliances. Now the goal is to develop a concept of ambiguous commitment that can be modeled theoretically and tested empirically. A number of scholars, including Fearon (1997), Zagare and Kilgour (2006), and Snyder (1984, 1997), have identified specific mechanisms through which a third-party state might form an ambiguous commitment. In the paragraphs that follow, I summarize their contributions and

consider their suitability as a general framework for ambiguous commitments. Ultimately, however, I conclude that although these models are an important point of departure, they are incomplete, as they either lack theoretical coherence or fail to encompass the full spectrum of commitments in the historical record.

Fearon (1997) reflects on the "puzzle" of "partial commitments," which are frequently observed in practice but do not obtain in equilibrium in his theory. He seems to have two types of mechanisms in mind. In one, a defender takes a costly action to signal commitment, but then reneges. Weak defenders might use this mechanism to bluff others into believing their commitment, but ultimately choose to back down in the face of a challenge they do not think they can win. The second mechanism Fearon mentions is a "half-hearted" signal in which a defender incurs a small cost that is less than the value needed to communicate a credible commitment. This mechanism is used by defenders that intend to assist the protégé even without the signal but that use the small cost to bluff an ambiguous commitment in order to restrain the protégé.

The first kind of bluffing is problematic not only because it is not equilibrium behavior in Fearon's model but also because, empirically, alliance members tend to keep their commitments (Leeds 2003a; Leeds, Long, and Mitchell 2000). The second mechanism – feigning weakness – does not obtain in equilibrium in Fearon's model because the challenger will infer that anything less than full commitment conveys unwillingness to intervene. This logic echoes Schelling's problem with the United States' use of ambiguity in the Quemoy crisis: "Any loopholes the threatening party leaves himself, if they are visible to the threatened party, weaken the visible commitment and hence reduce the credibility of the threat. (An example may be the ambiguous treatment of Quemoy in the Formosa Resolution and Treaty)" (1960, 40).

Fearon's and Schelling's observations that leaders, in fact, select ambiguous commitments despite the seeming irrationality of such choices result because they do not incorporate moral hazard into their theories. Weak signals based on partial or incomplete payments of costs are indeed irrational when the model does not take into account the protégé's response to the third party's signal. However, it is possible to imagine how a weak signal might be sustained in equilibrium if a protégé can be influenced by the signal and the adversary knows this.

Although the theory developed in this book fully considers moral hazard, the notion of ambiguity offered here steers clear of weak commitments or feints of weakness. The primary reason is that it is a challenge to identify such weak commitments in practice. Few, if any, alliances specify a level of military support that is obviously too low to be credible. Furthermore, although it may be that forming a firm and credible commitment does not entail high sunk costs or hands-tying, it is difficult to know this without also being able to identify empirically what the "critical value" of credibility is – in other words, the value of the cost that must be paid to make the commitment credible in the eyes of both the protégé and the adversary.

Moreover, observed diplomatic signals of weakness might result from other factors unrelated to moral hazard, such as the defender feeling torn between conflicting domestic and international pressures. For example, Grey's weakness in 1914 could be attributed to his desire to respond to the growing international emergency in the Balkans while simultaneously fearing backlash from a domestic audience dead set on neutrality (Levy 1990; Zagare and Kilgour 2006). Weak signals naturally result from these kinds of conflicting pressures. Thomas Christensen (1996) explains that weak, limping financial support from the United States to the Chinese Nationalists in the waning months of the Chinese Civil War in the late 1940s stemmed from the Truman administration's desire to abandon the Nationalists because of the on-the-ground reality that the Nationalists would soon lose the war, while simultaneously appeasing the strong domestic China lobby back in Washington. We would, therefore, be mistaken if we attributed the weak U.S. support of the Nationalists at that time as a shrewd strategy to deter the Chinese Communists from attacking the Nationalists while restraining the Nationalists from attacking the Communists.

If the theory offered in this book does not identify an ambiguous commitment mechanism as bluffing, then a potentially promising alternative is a mixed strategy. This is the approach taken by Frank Zagare and Marc Kilgour (2006), who develop a theoretical model of conflict with moral hazard and show that an equilibrium exists in which the defender randomizes between defending and abandoning its protégé to maximize the deterrent effect on the adversary while also restraining the protégé. However, to induce a mixed strategy equilibrium, the theory hinges critically on an assumption that protégés possess options for punishing defenders for not defending. In particular, Zagare and Kilgour's theory gives protégés the option of realigning with another state, and defenders do not know whether the protégé is the type to punish them by realigning.

In framing the problem this way, Zagare and Kilgour criticize Crawford's (2003) claim that simultaneously achieving deterrence and restraint is possible only when the protégé does not have outside alignment options. Zagare and Kilgour show that if the protégé has outside alignment options, the defender's deterrence threat against the adversary is most effective when the protégé can, in fact, also threaten to realign if the defender does not protect it. However, building a theory on the availability of outside alignment options, while making it attractive in examining the tractability of one explanation for ambiguous commitments, also makes it overly restrictive as a general framework for analyzing the problem of deterrence versus restraint; mixed strategies in the theory depend critically on the assumption that the defender cares about the threat of realignment, and it is uncertain about whether the protégé is of the type that would actually carry out the threat. This does not appear to be a natural way of thinking about many historical examples. In the Quemoy crisis, for instance, the United States did not worry that Chiang might realign with another state. Furthermore, in today's Taiwan Strait crisis, the United States is

the only prospective defender for Taiwan and, should the United States abandon it, Taiwan does not have other means by which it can punish the United States. The goal of the present study is to develop a concept of ambiguity that captures the idea that the defender's commitment generates probabilistic uncertainty, without also being dependent on the restrictive conditions necessary to induce mixed strategies in equilibrium.

Another way to characterize an ambiguous commitment is to think about the defender's strategy as an attempt to send simultaneous yet conflicting signals to the adversary and protégé. In Snyder's account (1984), deterrence and restraint can best be achieved by a strategy that instills and maintains opposite and pessimistic beliefs about the third-party defender's response in the event of war – the adversary should believe the defender will respond, and the protégé should believe that the defender is unlikely to. Instilling conflicting beliefs is a complicated diplomatic feat requiring at least two separate – and, preferably, private – messages. This was, in fact, Eisenhower's move in the Quemoy crisis. Not only was the public language of the formal treaty ambiguous, but the United States also insisted on the exchange of secret treaty notes with the Chinese Nationalists. Those notes clearly stipulated that the United States would use its discretion to determine whether to assist the Nationalists. Meanwhile, in an effort to send a conflicting message to Mao Zedong, both Dulles and Eisenhower boasted cavalierly about how the United States would not hesitate to use nuclear weapons against Communist China.

Identifying conflicting signals presents both theoretical and empirical challenges. Empirically, most signals of this kind are likely to be made in secret; others, made in public, would need to be distinguished from the noise of diplomatic exchanges. Given this research hurdle, it is worthwhile to take another approach. On the theoretical front, it is not clear whether a strategy of fooling both sides could be sustained in equilibrium. A commitment to issue conflicting signals might well unravel because the protégé and the adversary both recognize the third party's problem of simultaneously delivering on its conflicting obligations when the commitment is challenged.

Nevertheless, we might reasonably assume that if a single, observable ambiguous signal can be shown in theory to satisfy a defender's dual demands of deterrence and restraint, then multiple directional signals tailored both to deter and to restrain would also achieve the same objective, if such a strategy can be sustained in equilibrium. Indeed, it may often be the case that a defender will issue an ambiguous public commitment, which by itself is sufficient to satisfy the deterrence-versus-restraint dilemma, and then reinforce the message with secret but conflicting statements to the disputants. Therefore, for theoretical purposes, it is sufficient to demonstrate the conditions under which a single ambiguous instrument will be preferred to transparent and unambiguous alternatives in order to balance deterrence and moral hazard. Following this approach facilitates the empirical analysis because the comparative statics on the observable signal are testable, whereas directional private signals are not

(unless the researcher happens to be privy to the secret communications, as in the case of the Quemoy crisis). This approach can also be used in a qualitative analysis of a specific case.

As a result, the characterization of ambiguity developed in this book is a single public *ex ante* probabilistic commitment to perhaps intervene. This mechanism is conceptually distinct from feints of weakness, mixed strategies, and conflicting signals described earlier. In making a probabilistic commitment, the defender makes a promise, as credible as any transparent and unconditional signal, that with some probability it will intervene. In other words, leaders can design contracts to specifically stipulate that third parties might or might not intervene when there is war. In so doing, a defender may condition its response on or delegate the decision to intervene to some factor or decision-making mechanism external to the tripartite game. In practice, this might look like a promise that explicitly conditions the transfer of military assistance on the realization of some random process beyond the control of the third party and protégé. For example, the third party might benefit if it can commit to intervene with a 25 percent chance. It could, therefore, promise to intervene if two randomly selected world leaders are taller than a third randomly selected leader. In practice, many alliance commitments delegate the ultimate decision to intervene to some external decision-making body, and it is not known at the time the agreement is formed what the decision will be.

Why Ambiguity?

The theory in the book demonstrates that the third-party defender's choice of commitment mechanism to best satisfy the competing demands of deterrence and restraint depends on four main factors: the defender's power relative to the protégé and the adversary; the defender's preference for the protégé's security; the defender's preferences for how disputes involving its protégé are settled; and the observability of the hostile actions leading to war between the protégé and the adversary. Moral hazard is less of an issue if the defender values the protégé's security and shares the protégé's preferences for the settlement of the dispute. Divergent preferences, on the other hand, lead the defender to be vigilant about incorporating into the commitment mechanism specific conditions or ambiguity to reduce the moral hazard that occurs from an unconditional commitment. Powerful defenders are especially cautious because the size of the moral hazard distortion is correlated with how much the protégé expects to benefit from the third-party defender's assistance. One might think that a powerful third party could simply write a contract that specifies a promise to transfer only a limited amount of military assistance to the protégé when the *casus foederis* of the alliance contract has been triggered. Such a commitment, however, may not be credible if the defender has strong incentives to prevail in the war once it chooses to intervene. The inability to limit transfers in war complicates the contracting problem, not only because the defender cannot credibly