International Crises and Domestic Politics

ALASTAIR SMITH Yale University

Audience costs enable leaders to make credible commitments and to communicate their intentions to their adversaries during a crisis. I explain audience costs by simultaneously modeling crisis behavior and the domestic reelection process. I assume that a leader's ability influences the outcome of a crisis. As such, voters use outcomes as a signal of their leaders' quality. Leaders have incentives to make statements that deter their enemies abroad, since these statements also enhance their standing at home. Yet, such “cheap talk” foreign policy declarations are only credible when leaders suffer domestically if they fail to fulfill their commitments. In equilibrium, false promises are only made by the least competent types of leaders. Leaders that break their promises suffer electorally. Because initial domestic conditions and institutional arrangements affect the vulnerability of leaders to these domestic costs, such factors influence the credibility of policy declarations and, therefore, the crisis outcome.

In 1990, prior to the Gulf War, President Bush made explicit foreign policy statements regarding Iraq’s invasion of Kuwait. In short, he threatened Saddam Hussein. When this failed to cow Hussein into withdrawing, Bush followed through on his threats and authorized force to remove the Iraqi army from Kuwait. By comparison, U.S. policy toward Bosnia was less direct. Both presidents Bush and Clinton adopted vague, ambiguous policies toward the Bosnian crisis, and U.S. intervention was limited. Potentially, Clinton could have used direct military force to lift the siege of Sarajevo, but he neither pursued this strategy nor threatened to do so. Sometimes threats carry the weight of direct action. For example, Clinton outlined the contingent circumstances under which he would authorize military intervention in Haiti; just before the announced deadline, the Haitian military junta relinquished power, alleviating the need for military force.

When leaders make threats, these statements are often taken as meaningful. Yet, leaders are not bound by these statements in any formal, legal, or moral sense: Threats are “cheap talk.” When statements are nonbinding, why should other nations believe them? A common explanation is that leaders, particularly in democratic systems, pay “audience costs” if they fail to follow through on their commitments (Fearon 1994a, 1997; Martin 1993). Hence, having made belligerent statements and fearing the domestic consequences of reneging, leaders carry out their threats. Audience costs offer a mechanism via which leaders can credibly commit.

The concept of audience costs is problematic, however. It requires that citizens remove leaders who renge on their promises. Yet, if citizens are happy with their leader overall, then why should they throw her out of office simply because she fails to carry out one particular threat? Indeed, they may be delighted that she avoided foreign entanglement once she real-ized that she was unable to influence the situation. In the literature, audience costs are simply assumed to exist. I provide microfoundations for why citizens want to remove leaders who fail to honor their commitments. In addition to explaining why audience costs work, I predict how domestic conditions and political institutions affect the extent to which these costs bind leaders and, hence, the extent to which threats influence crisis behavior.

Information plays a key role in answering the audience cost puzzle. Leaders vary in quality, and this is reflected in their ability to perform during crises. Talented leaders, with the appropriate skills, are more likely to be successful during a crisis than their less competent peers. Since ability affects outcomes, voters want to remove incompetent leaders, but they can only perceive leaders’ ability through actions and performance. Thus, when choosing foreign policy, leaders must consider not only how it influences foreign rivals but also how it will be received at home. All else equal, leaders who expect to perform poorly in conflict avoid foreign entanglement. This is true for two reasons. The first is obvious: Poor performance is inherently undesirable. The second is that revealing low competence harms a leader’s electoral prospects. This is particularly important in lopsided conflicts, when the winner may not be in doubt, but efficiency reveals a leader’s ability.1

In this setting, reneging on a commitment is a sign of incompetence, which the voters punish electorally. To see why, suppose that audience costs do exist and that voters do indeed punish leaders who fail to follow through on threats. For competent leaders, carrying out threats imposes few costs. They expect to do well in the crisis, and it provides them with an opportunity to reveal their abilities to the electorate. For less competent leaders, having their bluff called leaves them between a rock and a hard place. If they renge on their commitment, then they are punished domestically, which means that many of them engage in conflicts they would prefer to avoid. For the least able leaders, however, conflict is impractical. Since their poor performance will be electorally punished anyway,

1 The arguments do not require that the marginal effect of ability on outcomes be large.

Alastair Smith is an Assistant Professor of Political Science, Yale University, New Haven, CT 06520-8301.

An earlier version of this paper was presented at the Peace Science Society meeting, Columbus, Ohio, October 13–6, 1995. I thank the participants at that meeting: John Ginkel, Andrew Martin, Fiona McGillivray, and James Morrow; and several anonymous referees for their useful comments.
they have no incentive to intervene, and they renege on their promises. Of course, this provides the justification for the electorate to punish them for failing to honor their commitments. Audience costs are endogenous. Since it is the least competent leaders who renege, broken commitments are a sign of incompetence which the voters punish.

This article ties together crisis behavior and domestic politics, simultaneously addressing how the latter shapes foreign policy and how crisis outcomes affect domestic political survival. In doing so, it produces a theoretical explanation of audience costs. Before presenting the full model, I consider different aspects of the puzzle. I start by examining a stylized model of extended deterrence and use it to explain the incentives that leaders face during international disputes. The conflictual nature of crises hampers communication between nations and leads to credibility problems. I explain why, in order to make threats or credible commitments, leaders must be able to send costly signals. Hence, intrinsically costless policy statements are informative only because of their domestic political implications. In this context, foreign policy choices have both international consequences—influencing the outcome of crises—and domestic consequences that affect voting decisions. After examining the links between domestic and international politics, I formally model foreign policy decision making within international crises against the backdrop of domestic politics. In doing so, I provide an explanation for the international/domestic nexus.

INTERNATIONAL CRISSES

I consider the case of extended deterrence among three nations (Huth 1988). I focus on this potential war scenario, but the incentives this type of event creates could apply equally well to less intense forms of conflict. In the stylized model, nation A is in a dispute with nations B and C over some issue. Suppose that nations B and C are initially content with the status quo, but nation A desires change. Nation A can attack nation B. If nation B acquiesces, then nation A gets its desired policy. If nation B resists, then the outcome depends upon which nation prevails in the conflict. I model the outcome of war as a lottery. Once a war starts, nation C can choose to intervene.

I illustrate the extensive form of the international crisis game in Figure 1. If A does not attack, then the status quo prevails. If A attacks and B surrenders, the outcome is acquiescence. Once A attacks and B resists, war occurs. The war becomes multilateral if C intervenes but remains bilateral if C stays neutral.

Each nation wants its preferred issue position to prevail. Yet, nations also want to avoid fighting. Conditional on having attacked, A’s preferences, from best to worst, are acquiescence, bilateral war, and multilateral war. When B surrenders, A is certain of attaining its goal. The more resistance A encounters, however, the smaller the chance that it wins. As B becomes likely to resist and C becomes likely to intervene, attacking becomes less attractive. While it is possible that A will attack knowing that this will lead to a multilateral war, the probability that A attacks is much higher if A believes B will surrender.

Both B and C want the status quo to prevail. Yet, if A attacks, then B must decide whether to fight or acquiesce. B is more likely to win any war if it receives allied intervention. Hence, B prefers a multilateral to a bilateral war, so its resistance depends in part on its expectations about C. Although C’s most preferred outcome is the status quo, it is A that chooses whether to disrupt it. C only gets to choose between intervening and remaining neutral. The decision to intervene depends upon whether the marginal effect of intervention on the probability of victory overcomes the costs of fighting. Intervention allows C to influence directly the crisis outcome, but C also can indirectly influence the course of events. Expectations about whether C will intervene affect whether A attacks and whether B resists. Therefore, if C can simply alter the expectations of A and B, then it can drastically alter the course of events.

CREDIBLE COMMUNICATIONS IN DETERRENT SITUATIONS

The international crisis game illustrates deterrence. Each nation’s decision depends upon what it believes the other nations will do. For example, the more likely C is to intervene, the more likely B is to resist. The more likely B is to resist, the less likely A is to attack.

C would like to deter A from attacking B and can do so by convincing A that it is prepared to intervene. As in the standard deterrence problem, however, convincing A is difficult. Suppose that C’s leader releases a foreign policy statement saying that she will intervene if A attacks. If A believes this message and is deterred,
then C benefits because the status quo is maintained. Yet, C benefits whether or not the policy statement is true. Providing A believes C’s claims, C rarely has to intervene, because A is unlikely to attack in the first place. Therefore, independent of what C really intends to do, if A is deterred by threats, then C should threaten to intervene. Unfortunately, this creates a credibility problem: C’s messages are unrelated to her intentions. A should ignore threats because they are vacuous, containing no information about whether C will really intervene.

In general, if A’s response to any foreign policy statement favors C, then C always sends the intervention message, whether or not it is true. But the message should have no effect on A, since it does not reflect anything about C’s likely decision. This creates something of a paradox. Theoretically, neither nations nor domestic electorates have any reason to believe or respond to any foreign policy statements. Yet, empirically, threats and ultimatums sometimes work and announcements of policy often affect public opinion. To rectify this discrepancy, I explicitly model the role of domestic politics within international crises.

In the context of game theory, foreign policy statements that have no intrinsic cost are commonly termed “cheap talk.” Jervis (1970) calls such costless messages signals. I show, as Jervis claims, that these signals can be informative, because they affect the beliefs, and hence the behavior, of domestic actors. Fearon (1994a, 1994b, 1997) correctly claims that in the international context alone, messages must be costly to be informative. Yet, foreign policy is not insulated from domestic concerns. Leaders simultaneously act at both the international and the domestic level. Providing that messages are costly in one setting, they can be informative in the other. Fearon assumes that there are exogenous domestic audience costs associated with foreign policy statements, and it is these which enable leaders to make credible threats. Specifically, Fearon assumes that leaders pay a fixed cost in terms of domestic support if they fail to follow through on their policy pledges. Why should voters want to punish leaders who bluff? Fearon assumes the existence of such costs but does not explain how they arise. In the international context, a leader who bluff may well be acting in her nation’s best interest. Suppose a leader, having failed to influence a crisis with threats, resists becoming embroiled in a war. Had her bluff succeeded, the nation as a whole would have benefited. Why should voters punish their leader when she believed that the war was not in her nation’s interest? Surely the voters prefer a leader that remains neutral when the war is not worthwhile. Under these circumstances, it seems inappropriate for them to punish leaders who fail to carry out a single threat.

By simultaneously modeling crisis behavior and domestic elections, I offer a theoretical explanation for audience costs. The key is not that voters are unhappy with their leader’s handling of the current crisis; rather, the failure to carry out threats reveals information about a leader’s ability. In other words, voters are content that their leader kept them out of the conflict, but in the future they want another leader. In addition to providing a microfoundation for audience costs, the model generates testable hypotheses relating domestic politics and foreign policy choice. Before presenting the model, I summarize some of the linkages between internal and external politics.

DOMESTIC POLITICS AND CRISIS BEHAVIOR

Political leaders have two audiences: domestic and international (Putnam 1988). Many scholars have observed that foreign policy and domestic politics are related (Russett and Graham 1989 summarize much of this literature). The more interesting issue is the direction of causality. Does the outcome of international events affect domestic survival, or does domestic politics drive the formation of foreign policy?

Studies of both voting decisions and public opinion conclude that foreign policy performance affects domestic political support (Aldrich, Sullivan, and Borgida 1989; Hurwitz and Pfeffley 1987; Mueller 1973; Ninic and Hinckley 1991). Although some of this research concludes that the domestic economy rather than foreign affairs is more important in explaining support, none finds international events unimportant. Furthermore, their influence is not limited to democratic systems. Bueno de Mesquita and Siverson (1995; also see Bueno de Mesquita, Siverson, and Woller 1992) find that, for all regime types, international outcomes affect domestic survival. For example, leaders who lose wars are more likely to be removed than successful leaders.

Other scholars suggest that the causal relationship runs in the opposite direction. For example, both Ostrom and Job (1986) and James and Oneal (1991) find that domestic political factors are more important in explaining foreign policy choice than are international factors. Regardless of an independent effect, the role of international events in domestic survival suggests that domestic politics should influence policy choice. Leaders who want to stay in office should select policies that help them achieve that goal. The choice depends upon domestic support and political institutions.

The diversionary war hypothesis suggests that leaders with domestic problems who anticipate being removed often undertake adventurous foreign policies that they would not otherwise attempt (Downs and Roche 1993; Hess and Orphanides 1995; Levy 1989; Richards et al. 1993; Smith 1996a). Since the model I present suggests similar results, it is worthwhile to examine the logic of this argument. Leaders often pursue risky policies abroad in order to divert attention from their domestic failures. Such policies may succeed, but often they do not. Since failure at the international level often results in removal, why would leaders take the chance? They may anticipate being removed anyway, so a policy failure would leave them.

---

4 Sartori (1996) suggests that reputational issues generate these costs.
no worse off, whereas success could restore their fortunes. Leaders who enjoy popular support do not enact such policies, not wanting to jeopardize their survival.

The diversionary hypothesis implies that domestic political conditions drive the formation of foreign policy, but the explanation for this linkage requires a more complex understanding of the relationship. It is the domestic consequences of international outcomes that drive leaders to select particular policies. This suggests that causality does not run in a single direction. In my model, in addition to the common goals of the nation, leaders are motivated to select foreign policies because of their domestic consequences. The interaction of policy choice and leader evaluation is embedded within the context of an international dispute.

While it is important never to forget policy differences, within the international context voters typically care about their leader’s ability. The outcome of international crises often resemble public goods. When a nation gains control of a resource or a piece of territory, this typically benefits the country as a whole. Therefore, the problem for the voters is not to find a leader with the same preference as themselves: All want their nation to succeed. Rather, at issue is finding a competent leader who can obtain the outcomes desired by voters. At least on international issues, I assume that voters choose between leaders on the basis of ability rather than policy differences.

In common with much of the literature, I assume that leaders differ in ability (Hess and Orphanides 1995; Richards et al. 1993; Smith 1996a). Some are competent, able to choose the right combination of forces, appoint skilled generals, adopt appropriate strategic plans, and persuade their allies to support them. Others are less adept. On average, competent leaders are more likely to be successful during an international crisis than are their less competent peers. More specifically, I assume that once a nation joins a war, the more competent its leader, the greater the probability that it will win. This is not to say that a leader’s ability is the only determining factor; obviously, as in the numerical examples I present, such structural differences as military strength may have a dominating influence. The assumption I make is that, at the margin, the quality of leadership affects who wins.

**THE MODEL**

There are four players in the model: the political leaders in nations A, B, and C and the domestic electorate in C. When I refer to A, B, and C, I mean the leaders as well as the states. The game has two stages, the international crisis phase and the domestic electoral phase. Figure 1 showed the international crisis stage. C sends a costless message indicating her foreign policy. Having observed this message, A decides whether to attack, B decides whether to resist, and C decides whether to intervene. Next, in the domestic electoral phase, the voters in nation C, having observed their leader’s message and the outcome of the international crisis game, decide whether to retain their leader.

Of particular substantive interest is how each nation behaves after observing a particular foreign policy message. I define the probability that A attacks after observing message m as ω(m). The probability of B retaliating after message m is β(m). Given message m, C intervenes with probability γ(m).

If a war occurs, then the expected outcome depends at least partly on the competence of the leaders involved. The more competent a nation’s leadership, the greater the likelihood the nation will be victorious. Let θa represent the competence of A, θb represent B’s competence, and θc represent C’s competence. Let Θ = (θa, θb, θc) be the competence of the leaders in each of the states. The probability that B wins a bilateral war is q(Θ). As B becomes more competent, q(Θ) increases. However, B is less likely to win if A is competent: q(Θ) is decreasing in θa. C, by intervening, increases the probability that B wins. The probability that B wins a multilateral war, p(Θ), is increasing in θb and θc and decreasing in θa. Hirschlief (1989) provides an interesting discussion of the form that q(·) and p(·) should take as a function of resources. Here I assume that resources are fixed and instead consider how q(·) and p(·) depend upon leader competence. In the numerical examples, for simplicity, I assume that p(Θ) and q(Θ) are simple linear functions of Θ. Specifically, p(Θ) = (θb + θc - θa)/6 + 0.6, and q(Θ) = (θb - θa)/6 + 0.55. This means that the most competent type of leader is about 17% more likely to be victorious than the least competent leader.

I assume that competence ranges between 0 and 1. Each leader knows her own ability but is uncertain of that of the others. In terms of incomplete information games, a leader’s competence is her type (Harsanyi 1967–68). Let μi(θi) represent the beliefs of the other players about the competence of the leader in nation i. For convenience, I assume that initial beliefs are uniformly distributed over the unit interval.

The payoffs from the international crisis game are summarized in Table 1. The most preferred outcome of B and C is the status quo. If B acquiesces, then B and

<table>
<thead>
<tr>
<th>Nation</th>
<th>Status Quo</th>
<th>Acquiescence</th>
<th>Bilateral War</th>
<th>Multilateral War</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0</td>
<td>1</td>
<td>1 - q(Θ) - ka</td>
<td>1 - p(Θ) - ka</td>
</tr>
<tr>
<td>B</td>
<td>1</td>
<td>0</td>
<td>q(Θ) - kb</td>
<td>p(Θ) - kb</td>
</tr>
<tr>
<td>C</td>
<td>1</td>
<td>0</td>
<td>q(Θ)</td>
<td>p(Θ) - kc</td>
</tr>
</tbody>
</table>
C receive a payoff of 0. If a war occurs, then their payoffs are the probability of victory: \( p(\Theta) \) if the war is unilateral or \( q(\Theta) \) if the war is bilateral, minus the cost \( (k) \) of fighting \( (k_b \) and \( k_r \), respectively). A receives a payoff of \( 0 \) for the status quo; its most preferred outcome is acquiescence. If A becomes involved in a war, then its payoff is its probability of victory: \((1 - p(\Theta))\) for a unilateral war or \((1 - q(\Theta))\) for a bilateral war, minus the cost of fighting, \( k_r \). In addition to caring about international outcomes, C receives a payoff of \( \Psi \) if she is reelected in the domestic electoral phase.

**Elections**

The election that follows the international crisis provides voters in C the opportunity to replace their leader. I assume that they evaluate her on two dimensions. First, they care about competence. Since it affects performance in future international events, the more competent they believe their leader to be, the more likely they are to retain her. Second, voters care about how their leader has performed on all other issues. Rather than explicitly modeling a leader’s performance on these other policy dimensions, I assume that her ability can be represented by a single parameter, \( bias \).

The bias parameter, which is analogous to Hess and Orphanides’s (1995) domestic handling ability characteristic, reflects a leader’s performance on issues other than the international dimension. It thus reflects the extent to which voters are predisposed to reelect the incumbent. If, for example, C has performed exceptionally well during her time in office (bias \( \gg 0 \)), then the voters are likely to retain her, independent of how well she performs on foreign policy. The voters are indifferent about retaining a leader with a mediocre record (bias \( \approx 0 \)), so their assessment of her international performance will have a large effect on whether she remains in office. When the electorate is predisposed against a leader (bias \( \ll 0 \)), she is unlikely to survive in office, regardless of foreign policy performance.

I assume that voters are rationally retrospective, using the information they have learned about a leader’s performance in deciding whether to retain her.\(^5\) When the election occurs, the voters compare their current leader to the challenger. The greater their predisposition (bias) toward the incumbent and the more competent they think she is, the more likely they are to retain her. Having seen the foreign policy message, \( m \), and the international outcome, \( z \), suppose the voters believe that the expected competence of C is \( E[\theta_c|m, z] \). Given these beliefs, the probability they will reelect the incumbent is \( \Phi(E[\theta_c|m, z], bias) \), where \( \Phi(\cdot) \) is a strictly increasing function of \( E[\theta_c|m, z] \) and bias. This function reflects the greater difficulty for a challenger to defeat popular, competent leaders than unpopular, incompetent ones. For the numerical examples, I assume that \( \Phi(\cdot) \) is the cumulative density function of a standard normal distribution evaluated at \((E[\theta_c|m, z] - 0.5 + bias)\). Thus, if bias = 0, then, ex ante, the probability of reelection is 50%. For brevity, I drop the bias term from the notation when it is not essential.

**Voter Beliefs**

Before the international crisis game, the voters’ prior beliefs about \( C \), \( \mu_c(\theta_c) \), are uniform. Then the voters learn about C’s type by her choices and performance. Given C’s strategy, Bayes’s rule defines the voters’ posterior beliefs about C’s type. For example, if only extremely competent types send message \( m' \), then upon seeing the message \( m' \), the voters (and other nations) believe that \( C \) is extremely competent. Alternatively, if all types send the same message, then the voters cannot update their beliefs upon observing the message.

Voters also learn about C’s type if she engages in a war. Since C’s type determines her ability to fight, I assume that C’s competence is fully revealed if she becomes involved in a war. This is the updating rule used by Hess and Orphanides (1995), and it is used to calculate the numerical examples. An alternative and equally realistic assumption, however, is that the voters observe C’s success during the war but do not observe her competence directly. Since voters know that more competent leaders are more likely to win, they can update their beliefs from the outcome of the war (the updating rule used in Richards et al. [1993] and Smith [1996a]). In this scenario, C’s incentives remain as described, although the numerical calculations will differ. The main propositions are not sensitive to this choice of assumption.

**RESULTS**

I discuss the results in two sections. In the first, crisis behavior in the absence of credible threats is examined. I explain how the domestic conditions in C affect the behavior of all nations involved in the crisis. In addition to generating useful predictions about how domestic political conditions influence foreign policy, these results lay the groundwork for explaining how audience costs are generated. In the second section, I discuss why foreign policy statements are credible and why domestic electorates punish leaders who renege on their threats. I explain the intuition behind the results and illustrate their implications using several numerical examples. The technical results, proofs of the proposition, and equilibrium refinements are discussed fully in the Appendix.

---

\(^5\) See Alesina, Londregan, and Rosenthal 1993; Ferejohn 1986; Persson and Tabellini 1990; Rogoff and Silbert 1988; and Smith 1996b for discussions about the differences between retrospective and prospective voters. I assume that international competence and domestic performance are independent.
Crisis Behavior in the Absence of Credible Threats: Babbling Equilibria

Consider the case in which foreign policy statements are uninformative. Suppose that, independent of her type, C randomly picks a foreign policy message. Such messages are typically referred to as babbles. Since each type is equally likely to send any particular message, neither foreign nations nor domestic audiences can update their beliefs about C's competence. Since messages are uninformative, all being identically interpreted, C is indifferent about which message to send. I characterize crisis behavior in babbling equilibria as follows.

Crisis Behavior. Competent leaders expect to perform well if they become involved in a war. Therefore, competent leaders are likely to behave aggressively and become involved in war (Banks 1990). The greater C's competence, the larger her marginal effect on the outcome of the war, and the more likely her assessment that intervention is worthwhile. Similarly, competent types of A and B expect to perform well should a war occur. Hence, when B is competent, B is more likely to resist; when A is competent, A is more likely to attack. Therefore, competent A types attack, competent B types retaliate, and competent C types intervene.

Proposition 1: For any beliefs about C's type \( \mu_c(\theta|m) \), the behavior of nations A, B, and C can be characterized by a unique triple: \((\theta^*_a(m), \theta^*_b(m), \theta^*_c(m))\). A only attacks if its type is greater than \( \theta^*_a(m) \), B only retaliates if its type is greater than \( \theta^*_b(m) \), and C only intervenes if its type is greater than \( \theta^*_c(m) \). Having observed \( m \), the probability that A attacks is \( \alpha(m) = 1 - \theta^*_a(m) \), the probability that B retaliates is \( \beta(m) = 1 - \theta^*_b(m) \), and the probability that C intervenes is \( \gamma(m) = \int_{\theta^*_c(m)} f_{\theta^*_c}(m) \mu_c(\theta|m) d\theta_c \).

Consider A's decision to attack. Given \( \theta^*_a(m) \) and \( \theta^*_b(m) \), A knows the probability that B retaliates and that C intervenes. In addition, A knows the distribution of competence for those types that resist and therefore knows how well he expects to do if he becomes involved in war. Anticipating how B and C will respond, each type, \( \theta^*_a \), calculates the expected payoff from attacking. The higher A's type, the larger this expected reward. \( \theta^*_a(m) \) is simply the type that is indifferent about attacking. Those types above \( \theta^*_a(m) \) attack, those types below \( \theta^*_a(m) \) do not. A and C structure their decisions in an analogous manner.

Crisis outcomes are informative; they tell us something about each leader's type. Since the more competent types of A attacks, if the crisis outcome is the status quo, then A must be a lower type. If he were not, then he would have attacked. When A fails to attack, observers should shift beliefs about A's type downward. Alternatively, their beliefs about A's type increase if A attacks.

Observers can similarly update beliefs about B and C given their actions. For example, if a war occurs and C fails to intervene, then C must be a lower type. If she were a higher type, then she would have intervened. Therefore, following a bilateral war, C is viewed as less competent than previously thought. Since voters care about the competence of their leader, they are less likely to retain one who fails to intervene. Next I focus on how this information affects voting decisions and drives leaders toward hawkish policies.

Domestic Politics Makes Leaders Hawkish

Proposition 2: Domestic politics makes C more likely to intervene than international factors alone would predict.

On the basis of international factors alone, C should only intervene if the marginal effect on the probability of victory outweighs the costs of fighting. Mathematically, C should only intervene if \( \bar{p}(\theta) - k_c \geq \bar{q}(\theta) \), where \( \bar{p}(\theta) = \int_{\theta^*_a(m)} f_{\theta^*_a}(m) p(\theta|m) d\theta_a d\theta_b \), the average probability of victory if C intervenes, and \( \bar{q}(\theta) = \int_{\theta^*_b(m)} f_{\theta^*_b}(m) q(\theta|m) d\theta_a d\theta_b \), the average probability of victory for B in a bilateral war. Domestic political incentives encourage C to intervene even when this condition is not met. Hence, C may intervene when the costs outweigh the expected benefits. Why? The answer lies in what the voters infer when a leader fails to intervene.

Suppose that leaders decide to intervene on the basis of international concerns only. Under these conditions, more competent leaders intervene, and less competent types remain neutral. Suppose that in a particular case, on the basis of international factors alone, only types with competence greater than \( \frac{1}{2} \) intervene. When the voters see the outcome of the crisis, what do they infer about their leader? If C intervenes, then the voters observe C's type directly, so they know exactly how competent she is. If C does not intervene, then the voters only know that C's competence is less than \( \frac{1}{2} \). Thus, the expected competence of types who do not intervene is \( \frac{1}{4} \).

Now suppose that C's type is \( \frac{1}{6} \). Since this is less than \( \frac{1}{2} \), on the basis of international considerations alone, she should not intervene. Yet, domestic considerations suggest otherwise. If C remains neutral, then the voters know that her type is less than \( \frac{1}{2} \). If she intervenes, then the voters will find out that her type is \( \frac{1}{6} \). The voters are more likely to retain a leader of competence \( \frac{1}{6} \) than one of expected competence \( \frac{1}{4} \). By intervening, C enhances her reelection prospects.

Domestic political incentives bias leaders toward hawkish strategies. The magnitude of this effect depends upon domestic conditions. Suppose that C is extremely popular, that is, bias is high, and she is likely...
to be reelected. She realizes that her international performance is unlikely to affect her chances in the next election. Since she expects to be reelected whatever her actions, her predisposition toward violence is small. Similarly, if $C$ is extremely unpopular, bias $\ll 0$, then international performance alone is unlikely to make a big difference in the electoral outcome. Again, leader $C$’s prejudice toward violence is small. Leaders have the largest predisposition toward violence when elections are anticipated to be close. When the voters are not heavily inclined toward a leader, their evaluation of international competence has a large effect on their decision. In the numerical example, I model the reelection decision using a cumulative normal density function. As is commonly observed in probit models, the marginal effect of foreign policy evaluation is greatest when, ex ante, the election is anticipated to be close. A numerical example is shown in Figure 2. The probability of intervention, $\gamma(m)$, is greatest when elections are close (bias $\approx 0$). As the marginal influence of international competence on an election declines (high or low bias), $C$ is less likely to intervene. This is consistent with empirical observations by Russett (1990; also see Morrow 1991) that mild but not severe recessions are associated with the use of force abroad.

This has interesting implications for the interpretation of foreign policy. To illustrate, consider an extremely unpopular leader (bias $\ll 0$) and one of average popularity (bias $\approx 0$). The former is unlikely to survive even with a foreign policy success, therefore, she has little domestic incentive to intervene and the probability of intervention is low. The latter is more likely to intervene, given the marginal effect of foreign policy on elections. How voters interpret nonintervention varies. In the latter case, many types intervene, and the types who fail to intervene must be the very lowest types. If $C$ fails to intervene, then the voters believe she has little competence. When the probability of intervention is low, however, failing to intervene carries less information; an unpopular leader who remains neutral may appear less competent to voters than previously, but the downward shift is small. How the electorate evaluates nonintervention depends upon initial domestic conditions.

The domestic context affects how a policy is interpreted, a leader’s choice of actions, whether a leader sends informative messages, and how those messages are viewed. This has implications for empirical studies of the relationship between domestic politics and foreign policy. The model indicates that the magnitude of
foreign policy effects on public opinion and reelection depends upon initial domestic conditions.

The more leaders care about reelection (increasing $\Psi$), the greater the probability that they will intervene. Policy-driven leaders are less inclined toward ill-advised foreign policy actions than leaders who care primarily about reelection. This has electoral cycle implications (Gaubatz 1991; Morrow 1991; Stoll 1984). As new elections approach, we might suspect that leaders become increasingly concerned about reelection and therefore hawkish late in their term (Downs and Rocke 1993; Hess and Orphanides 1995; Smith 1996a).

**Initiation, Retaliation, and Intervention**

**Proposition 3:** Intervention and initiation are negatively related. As the probability that $C$ intervenes increases, $A$ is less likely to attack.

Figure 2 illustrated the relationship between initiation, retaliation, and intervention with a numerical example. $C$’s propensity to intervene depends upon her internal political situation. When the voters’ evaluation of her international performance strongly influences the electoral outcomes ($\text{bias} \approx 0$), $C$ is likely to intervene. If $A$ attacks under these conditions, then $A$ should anticipate a multilateral conflict. Yet, if $C$ is popular ($\text{bias} \approx 1$), then $C$ is unlikely to intervene, and $A$ should anticipate a bilateral conflict. Since $A$ prefers a bilateral to a multilateral war, $A$ is more likely to attack when $C$ is extremely popular rather than when $C$ has mediocre support. Interestingly, $A$’s decision to initiate violence depends, at least in part, upon domestic conditions in another nation.

In general there is an ambiguous relationship between retaliation, $\beta$, and intervention, $\gamma$. Intervention can either increase or decrease the probability of retaliation, depending upon the parameters of the particular case. Two competing factors influence $B$’s decision. The likelihood of intervention means that $B$ anticipates more support in its fight against $A$, which encourages $B$ to retaliate. Yet, since intervention should deter $A$ from attacking, if $B$ is actually attacked, then it knows its aggressor is extremely competent and may decide to acquiesce. In Figure 2 intervention and retaliation are not positively related, the latter influence outweighing the former, but this is not a general result; under other conditions $\beta$ and $\gamma$ are positively related.

The significance of proposition 3 is that domestic conditions in nation $C$ affect foreign policy decisions in nations $A$ and $B$ (Morrow 1991). As the probability of intervention increases, the probability of initiation and retaliation declines. This creates a sample selection problem in testing how domestic affairs affect $C$’s foreign policy (Achen 1986; Fearon 1993; Morrow 1989; Smith 1996c). We can only observe $C$’s intervention decision when war occurs, but the probability of war depends upon what $C$ is likely to do should war actually occur. In terms of empirical testing, this is problematic and leads to incorrect statistical inference unless we control for the selection effect.

The discussion of uninformative equilibria shows that (1) domestic politics can make $C$ hawkish, (2) domestic politics in $C$ affect $A$ and $B$’s decisions, and (3) initial domestic conditions affect how voters interpret a leader’s actions. These properties remain true in informative equilibria.

**Making Threats Work**

In order for threats to work they must convey information about a leader’s intentions. When messages are costless to send, it is difficult for leaders to signal their type (Crawford and Sobel 1982). Yet, leaders can make credible threats in the international arena because of the domestic ramifications of failing to follow through. I show that if messages are informative, then the electorate wants to punish leaders who fail to honor their policy commitments. This generates audience costs and allows leaders to send informative messages. I explain the logic of the argument using an example. The properties of informative equilibria are formally characterized in the Appendix.

**An Informative Message Equilibrium.** In this equilibrium, competent leaders threaten to intervene; less competent leaders do not threaten. Should war occur, only those leaders who made threats intervene. Figure 3 shows a pictorial representation. More formally, there is a critical type $\theta^c$. Leaders who are more competent than this critical type ($\theta > \theta^c$) send message $m_1$ and intervene if a war occurs. Lower types ($\theta < \theta^c$) send message $m_2$ and remain neutral if a war occurs. Substantively, $m_1$ can be thought of as a threat to intervene. Since different types send different messages, the behavior of voters and foreign nations depends upon whether $C$ makes threats. Before explaining why $C$’s threats are credible, I examine how voters and foreign nations respond to each message.

If the voters observe a threat (message $m_1$), then they infer that $C$ is competent: Specifically, $\theta_c \geq \theta^c$. Therefore, the expected competence of $C$ is $(1 + \theta^c)/2$. Less competent types ($\theta_c < \theta^c$) do not make threats (message $m_2$). The expected competence of a type who sends $m_2$ is $\theta^c/2$. Since $(1 + \theta^c)/2 > 1/2 > \theta^c/2$, leaders receive an increase in electoral support if they threaten, but the electoral support for leaders who remain passive declines. In equilibrium, types who send message $m_2$ never intervene, so the voters never get the opportunity to update their beliefs about these types. The voters may get additional opportunities to learn about those types who do make threats. If $C$’s bluff is called and a war occurs, then the voters observe whether $C$ carries out her threats.
In equilibrium, only competent leaders make threats. These types also carry them out. If $C$ makes a threat, then $A$ knows that $C$ is a competent type ($\theta_c \geq \theta_c^*$) who will intervene ($\gamma(m_1) = 1$). If $C$ makes no threat, then $A$ knows that all wars will remain bilateral ($\gamma(m_2) = 0$). Since $A$’s probability of initiation declines as $C$ intervenes more often, threats deter $A$ from attacking ($\alpha(m_1) < \alpha(m_2)$). By making threats, $C$ benefits internationally because $A$ attacks less often.

There are clear advantages to making threats. $C$ benefits domestically, by increasing her chance of reelection, and internationally, by deterring $A$ from attacking. Why does every type not send $m_1$? Why are $C$’s threats credible? Despite the advantages of sending $m_1$, there is the disadvantage of audience costs associated with unfulfilled threats. Leaders are punished electorally by the voters if they fail to keep their promises. In equilibrium, all types who threaten to intervene, $m_1$, actually do so. What should the voters believe if $C$ threatens but does not intervene? Unfortunately, Bayes’s rule is undefined in these circumstances. Therefore, any beliefs are consistent with the sequential equilibrium concept. Standard refinements (discussed in the Appendix) suggest that the type with the largest incentive to remain neutral once having sent $m_1$ is the lowest type, $\theta_c = 0$. Using this restriction, the voters believe the expected competence of a type who makes threats (sends $m_1$) but does not intervene is zero: $E[\theta_c|m_1, \text{ Bilateral war}] = 0$. Leaders who fail to fulfill their foreign policy commitments have little chance of reelection. $E[\theta_c|m_1, \text{ Bilateral war}] = 0 < E[\theta_c|m_2] = \theta_c^*/2 < E[\theta_c|m_1] = (1 + \theta_c^*)/2$. This reduction in the probability of reelection is what Fearon (1994a) calls a domestic audience cost. It is a loss that leaders suffer when they do not honor their commitments. Yet, it is neither an exogenous fixed cost nor a mechanism to take the decision out of the leader’s control (Fearon 1997). Leaders choose to honor their commitments because of the electoral consequences of failing to do so. These costs are endogenously generated by equilibrium behavior.

Although threats deter $A$ from attacking, sometimes $C$’s bluff is called, and $C$ must choose whether to intervene. High types intervene because, being competent, they expect to perform well in the war, and it affords them an opportunity to show their competence to the voters. For less competent types, intervention is unattractive; they not only will perform badly but also will reveal their low competence to the voters. Yet, failing to intervene is also unattractive, since the voters will punish them electorally. Rather than risk this eventuality, low types send message $m_2$. For high types the cost of carrying out threats is low. For type $\theta_c^*$, the electoral and deterrence advantages of $m_1$ and the costs of having to intervene should a war occur exactly counterbalance the payoffs associated with sending $m_2$. 

631
and remaining neutral (type $\theta^2_C$ is indifferent between sending $m_1$ and $m_2$). Types below $\theta^2_C$ prefer to play safe, sending $m_2$, rather than risk sending $m_1$ and having their bluff called. Types above $\theta^2_C$ prefer to send $m_1$ because, being competent, they face lower costs for honoring their commitments.

Figure 4 illustrates a numerical example of the informative message equilibrium. In the example, $A$ always attacks unless it receives a threat, $m_1$. Upon observing a threat, $A$ believes that $C$ will intervene. $C$’s cheap talk partially deters $A$ from attacking because $C$’s threat is credible.

Leaders are able to send credible messages because they mortgage their domestic political survival on their ability to honor their commitments. Without the ability of the electorate to remove leaders, leaders cannot credibly communicate their intentions. Domestic conditions limit whether leaders can make credible statements. Absolute rulers, who cannot be removed, cannot communicate their intentions to other nations. Similarly, lame-duck presidents find it hard to make credible commitments. Without voter retaliation to keep them honest, their threats are less likely to be believed.

Since domestic politics enables leaders to communicate credibly, domestic circumstances affect which messages are sent and how these messages are interpreted. The electoral advantage of convincing the voters that you are competent depends upon your initial domestic prospects. Thus, domestic factors affect both the probability with which $C$ threatens and how other nations interpret these threats. Figure 4 showed that if $C$ is popular (bias $\approx 0.75$), and hence likely to get reelected, then fewer types consider it worthwhile to threaten (send $m_1$). Note the probability that $A$ will attack, given the threat $\alpha(m_1)$, declines as bias increases. As fewer types send the threat, the average competence of the types who intervene increases. Since $A$ has to combat a more competent type, it attacks less often when threatened. Interestingly, it is when the domestic incentives to send threats are small, and hence few types make threats, that threats have the largest deterrent effect.

**General Properties of Informative Message Equilibria.** The properties of informative message equilibria are characterized in the Appendix (lemmas 2, 3, and 4). The following statements summarize the substance of these lemmas. In general, the existence of informative message equilibria depends upon the ability of leaders to
mortgage their domestic political future. Equilibria can only be informative if war occurs with positive probability.

Without the possibility of war, leaders never face the prospect of having to honor their commitments. Under these circumstances, all types send any message that deters A, since they know that they will never have to make good on their promise. Since all types can send any message that benefits them, there is no useful communication. The possibility of war is necessary to keep leaders honest.

CONCLUSION

The model leads to two sets of results. First, it characterizes crisis behavior and shows how domestic conditions within a single nation affect foreign policy decisions by all nations involved in a crisis. Second, it shows how this linkage between domestic politics and foreign policy enables leaders to issue credible threats that deter adversaries. By explaining why voters punish leaders who fail to follow through on their threats, I generate a theory for the creation of audience costs. These costs make foreign policy statements meaningful.

Political leaders have two audiences, their domestic constituents and their foreign rivals. When forming foreign policies, leaders simultaneously balance these internal and external constraints. Hence, international events and domestic political survival are intrinsically linked, not through a simple unidirectional causal pathway, but via a series of strategic interactions at both the international and the domestic level.

A consequence of these linkages is that leaders adopt hawkish policies. Voters want to retain competent leaders and remove incompetent ones. Since competent leaders expect to perform better, they are more likely to intervene in international disputes. Given this, nonintervention signals lower competence. Leaders who want to retain office avoid being labelled incompetent by intervening. This desire to separate themselves from the low competence types creates a bias toward violent behavior. They enter wars in which, on the basis of international factors alone, they should not intervene.

The extent of this hawkish bias depends on domestic political circumstances. When the voters' evaluation of a leader's competence will have little effect on the election, leaders have little incentive to behave aggressively, and the bias toward violence is small. As a result, leaders who are either extremely secure or extremely insecure base their policies mainly on international considerations. For example, absolute rulers or lameduck presidents intervene on the basis of international conditions alone. When leaders anticipate close elections, their competence is an important issue, and the incentive to distinguish themselves make them extremely hawkish. This does not necessarily imply that they become involved in more wars. As domestic politics makes leaders more aggressive, it also deters other nations from initiating violence. As the probability of intervention increases, \( \gamma(m) \), the probability of initiation, \( \alpha(m) \), declines. So elections reduce the opportunities for intervention but increase a nation's propensity to intervene in each particular case. In general, the net effect of domestic politics on the occurrence of war is ambiguous; it depends upon the characteristics of each case.\(^6\)

The interaction of domestic and international politics enables leaders to make credible threats. Simple statements of foreign policy are informative because of the domestic electoral implications of not following through with an announced policy. The model explains how these audience costs arise. Since voters punish leaders who do not carry out their threats, leaders run a risk if they make threats. If their bluff is called and a war occurs, then they must decide whether to intervene, and failure to do so harms reelection prospects. A leader's competence determines the benefit of intervening. A highly competent leader may actually welcome the opportunity, since it allows her to demonstrate her superior abilities to the voters. Less competent leaders, although they want to avoid fighting, are also compelled to intervene to avoid being punished electorally. For the least competent leaders, intervention is not an option; their poor performance would provide yet more evidence of their low quality, and they will suffer the same electoral fate whatever they decide to do. This provides the justification for audience costs in the first place. If a leader fails to follow through on a threat, then the voters should infer that she is incompetent and want to replace her.

The model explains the origin of domestic audience costs. Threats deter foreign aggression through two mechanisms: commitment and information. Audience costs allow leaders to commit to a course of action that they otherwise would not undertake. Once a threat is made, a leaders faces a cost for backing down, and this cost makes intervention more attractive. Threats also work because they inform aggressors that their adversary is competent. Since competent leaders face the lowest costs for carrying out threats, they find the use of threats attractive. When an aggressor receives a threat, he knows that the leader not only is committed to carrying it out but also is competent. The prospect of becoming involved in a multilateral conflict against a competent adversary deters aggression.

When leaders make threats they jeopardize their domestic political prospects. The extent of their domestic vulnerability is associated with the degree of audience costs. This supports Fearon's (1994a) conjecture that electorally vulnerable democratic leaders

\(^6\) Other studies that observe this ambiguity include Fearon 1994a, Morrow 1994, and Smith 1995.
have larger audience costs than their autocratic rivals. At the extreme, when leaders have nothing to risk, they cannot commit to carrying out their threats. Therefore, their foreign policy announcements are uninformative.

Foreign policy statements convey information about a leader’s type and propensity to intervene. Such statements can deter foreign aggression because the pledge to intervene is credible. They also affect public opinion because they tell voters about the government’s level of competence. The exact interpretation of a foreign policy message and its effect on other nations and public opinion at home depends upon initial domestic conditions. Empirical assessments of foreign policy should consider the context in which a particular policy is taken.

**APPENDIX**

**Rigorous Description of the Model**

C announces a foreign policy message, \( m \in M \). Having observed this message, \( A \) chooses whether to attack (\( att \), \( \neg att \)). If \( A \) attacks, then \( B \) chooses whether to retaliate (\( ret \), \( \neg ret \)). If \( B \) retaliates, then \( C \) chooses whether to intervene (\( int \), \( \neg int \)). Let \( Z = \{MWAR, BIWAR, ACQ, SQ\} \) be the set of international outcomes. If \( A \) attacks, \( B \) retaliates, and \( C \) intervenes, then a multilateral war occurs, \( MWAR \). If \( A \) attacks, \( B \) retaliates, but \( C \) does not intervene, then a bilateral war occurs, \( BIWAR \). Acquiescence (\( ACQ \)) occurs if \( A \) attacks but \( B \) does not resist. Finally, the outcome is the status quo (\( SQ \)) if \( A \) does not attack. The payoffs from the international outcome, shown in Table 1, depend upon the competence of the nations.

The competence, or type, of the nations is \( \Theta = (\theta_a, \theta_b, \theta_c) \). Let \( \mu_k(\theta_i) \) be the prior probability density over \( \theta_i \). Although I assume that \( \theta_i \in \{a, b, c\} \) is distributed uniformly over the unit interval \( U(0, 1) \), all that is important for the following results is that the distribution over types is continuous and bounded away from zero for all types. Each nation knows its own type but only knows the distribution of the other nations’ competence. Competence affects the expected outcome of the war: \( p(\Theta) \rightarrow [0, 1] \) and \( q(\Theta) \rightarrow [0, 1] \). \( B \)'s probability of winning increases if \( C \) intervenes: \( p(\Theta) > q(\Theta) \). I assume that both \( p(\cdot) \) and \( q(\cdot) \) are differentiable in all components: Specifically, \( dp(\Theta)/d\theta_a < 0, dq(\Theta)/d\theta_a < 0, dp(\Theta)/d\theta_b > 0, dp(\Theta)/d\theta_b > 0, dq(\Theta)/d\theta_b < 0, \) and \( dq(\Theta)/d\theta_b < 0 \). The competence of any nation that becomes involved in a war is revealed.

\( C \)'s strategy is the pair \( (\sigma_c, \gamma) \), where \( \sigma_c : \Theta_c \times M \rightarrow [0, 1] \) and \( \gamma : \Theta_c \times M \rightarrow [0, 1] \). \( A \)'s strategy is \( \sigma_a : \Theta_a \times M \rightarrow [0, 1] \). \( B \)'s strategy is \( \sigma_b : \Theta_b \times M \rightarrow [0, 1] \). Thus, \( \sigma(b, \theta_c) \) is the probability that \( \theta_c \) sends message \( m \) when \( \theta_b \) sends message \( m \). \( \sigma(m, \theta_a) \) is the probability that \( \theta_a \) attacks having observed the message \( m \); \( \gamma(m, \theta_c) \) is the probability that \( \theta_c \) intervenes given message \( m \). Let the notation \( m(\theta_c) \) mean that \( \theta_c \) sends message \( m(\theta_c) \) with positive probability.

In addition to payoffs from international outcomes, the leadership of \( C \) receives a payoff of \( \Psi > 0 \) if reelected by the voters following the international crisis. The probability that the voters reelect \( C \), \( \Phi(E[\theta|m, z]) \), bias increases with the expected competence of \( C \) and the electorate’s bias toward \( C \) (Hinich 1978). The justification for this assumption is that, after the international crisis but before the election, the voters learn about the qualities of the alternative leadership. At the election, the voters choose the leadership they prefer. When the government makes its decisions it does not know what information the voters will receive. If the information about the opposition party is distributed normally, then when the leader makes her decision, the voters’ decision rule appears probabilistic. Once the information is revealed the actual voting decision is deterministic. Yet, prior to this revelation, voting appears probabilistic (Austen-Smith 1987; Smith 1996b).

The solution concept is sequential equilibrium (Kreps and Wilson 1982). This requires that the strategies of \( A, B \), and \( C \) are utility maximizing, given the strategy of the other players and their beliefs. The strategy of the voters, reecting \( C \) probabilistically, is utility maximizing by assumption. In addition, sequential equilibria require that all players’ beliefs are consistent with the Bayesian rule.

**Results**

**C's Decision to Intervene.** \( C \)'s expected utility for intervening is

\[
U_c(\int m) = \int_0^1 \int_0^1 \mu_a(\theta_a|att)\mu_b(\theta_b|ret)p(\Theta)d\theta_a d\theta_b
\]

\[
- k_\lambda + \Psi*\Phi(E[\theta_c|m, MWAR])
\]

where \( \mu_a(\theta_a|att) \) is posterior beliefs about \( A \)'s type, given that it attacks, and \( \mu_b(\theta_b|ret) \) is posterior beliefs about \( B \)'s type, given that it retaliates.

\( C \)'s expected utility for not intervening is

\[
U_c(\neg int m) = \int_0^1 \int_0^1 \mu_a(\theta_a|att)\mu_b(\theta_b|ret)q(\Theta)d\theta_a d\theta_b
\]

\[
+ \Psi*\Phi(E[\theta_c|m, BIWAR])
\]

Since \( p(\Theta) \) is increasing in \( \theta_c \), \( U_c(\int m) \) is strictly increasing in \( \theta_c \). \( U_c(\neg int \theta_c) \) is constant in \( \theta_c \). Define \( \theta^*(m) \) as the type indifferent about whether to intervene (if such a type does not exist, then set \( \theta^*(m) \) to the appropriate limiting value; that is, if all types prefer to intervene, then \( \theta^*(m) = 1 \); if all types prefer not to intervene, then \( \theta^*(m) = 0 \)). Thus, \( C \)'s equilibrium strategy is \( \sigma_c(\theta_c, m) = 1 \) if \( \theta_c \geq \theta^*(m) \), and \( \sigma_c(\theta_c, m) = 0 \) otherwise.

Let \( \gamma(m) \) represent the probability that \( C \) intervenes, given the message \( m \). Thus, \( \gamma(m) = \int_0^1 \mu_c(\theta_c|m)d\theta_c \), where \( \mu_c(\theta_c|m) \) is the posterior distribution of \( \theta_c \), given the message \( m \).

**B's Decision to Retaliate.** If \( B \) retaliates, then its expected utility, given its type, is

\[
U_r(\text{ret } \theta_i, m) = \int_0^1 \mu_r(\theta_r|att)\int_0^1 \mu_c(\theta_c|m)p(\Theta)d\theta_c
\]

\[
+ (1 - \gamma(m))q(\Theta) d\theta_a - k_b.
\]

If \( B \) does not retaliate, then it receives a payoff of \( U_r(\neg \text{ret } \theta_i, m) = 0 \).

Since \( p(\cdot) \) and \( q(\cdot) \) are increasing in \( \theta_i \), \( U_r(\text{ret } \theta_i, m) \) is strictly increasing in \( \theta_i \). Define \( \theta^*(m) \) as the type indifferent about whether to retaliate (if such a type does not exist, then
set $\theta^*_m(m)$ to the appropriate limit). Thus, $B$'s equilibrium strategy is $s_b(\theta^*_m, m) = 1$ if $\theta^*_m \geq \theta^*_m(m)$, and $s_b(\theta^*_m, m) = 0$ otherwise. Ex ante, the probability that $B$ retaliates, given the message $m$, is $\beta(m)$, where $\beta(m) = \int_{\theta^*_0(m)}^{1} \mu_b(\theta, m) d\theta = 1 - \theta^*_m(m)$.

**A's Decision to Attack.** If $A$ attacks, then its expected utility is

$$U_a(\text{att}|\theta^*_m, m) = 1 - \beta(m) + \int_{\theta^*_0(m)}^{1} \int_{\theta^*_0(m)}^{1} \mu_a(\theta, m) \left( p(\Theta) - k_c \right) d\theta_a d\theta_b + (1 - \gamma(m)) \int_{\theta^*_0(m)}^{1} \mu_a(\theta, m) \mu_a(\theta, |m) d\theta_a - k_c.$$ 

If $A$ does not attack, then its expected payoff is $U_a(-\text{att}|\theta^*_m, m) = 0$. Unless $\beta(m) = 0$, in which case all types attack, A's expected payoff from attacking is strictly increasing in $\theta^*_m$. Define $\theta^*_m(m)$ as the utility indifference type for whether to attack (if such a type does not exist, then set $\theta^*_m(m)$ to the appropriate limit). Thus, $A$'s equilibrium strategy is $s_a(\theta^*_m, m) = 1$ if $\theta^*_m \geq \theta^*_m(m)$, and $s_a(\theta^*_m, m) = 0$ otherwise. Ex ante, the probability that $A$ attacks, given the message $m$, is $\alpha(m)$, where $\alpha(m) = \int_{\theta^*_0(m)}^{1} \mu_a(\theta, m) d\theta_a = 1 - \theta^*_m(m)$.

**Proposition 1.** For any posterior beliefs about $C$'s type, given the message $m$, $\mu_a(\theta, m)$, there exists a unique triple: $(\theta^*_0(m), \theta^*_m(m), \theta^*_m(m))$. A only believes if its type is greater than $\theta^*_0(m)$, B only retaliates if its type is greater than $\theta^*_m(m)$, and C only intervenes if its type is greater than $\theta^*_m(m)$. Having observed $m$, the probability that $A$ attacks is $\alpha(m) = 1 - \theta^*_m(m)$, the probability that $B$ retaliates is $\beta(m) = 1 - \theta^*_m(m)$, and the probability that $C$ intervenes is $\gamma(m) = \int_{\theta^*_0(m)}^{1} \mu_a(\theta, m) d\theta_a$. (Proof follows directly from the above description of $A$, $B$, and $C$'s decisions.)

**Bayesian Updating and Out-of-Equilibrium Beliefs.** C's strategy, $\sigma_c(m, \theta)$, is the probability with which type $\theta$ sends message $m$. By Bayes's rule and initial uniform beliefs, the posterior beliefs about C's type are $\mu_c(\theta, m) = \sigma_c(m, \theta) / \int \sigma_c(m, \theta) d\theta$. The posterior beliefs on the equilibrium path can be parameterized, given the strategies above. If $A$ attacks, then $\mu_a(\theta, att) = 1/(1 - \theta^*_m(m))$ if $\theta_b \geq \theta^*_m(m)$, otherwise $\mu_a(\theta, att) = 0$. If $A$ does not attack, then $\mu_a(\theta, att) = 1/\theta^*_m(m)$ if $\theta_a < \theta^*_m(m)$, otherwise $\mu_a(\theta, att) = 0$. If $B$ retaliates, then $\mu_b(\theta, ret) = 1/(1 - \theta^*_m(m))$ if $\theta_b \geq \theta^*_m(m)$, otherwise $\mu_b(\theta, ret) = 0$. If B does not retaliate, then $\mu_b(\theta, ret) = 1/\theta^*_m(m)$ if $\theta_b < \theta^*_m(m)$, otherwise $\mu_b(\theta, ret) = 0$.

Given the beliefs about C's type, $\mu_c(\theta, m)$, if C intervenes, then $\mu_c(\theta, m, MUWAR) = \mu_c(\theta, m) / s_c(\theta, m)$ if $C$ does not intervene, then $\mu_c(\theta, m, BIWAR) = \mu_c(\theta, m) / (1 - s_c(\theta, m))$. Along the equilibrium path, the Bayesian rule defines beliefs, but the rule does not tell us what nature should believe if they observe an action that should not occur in equilibrium. The assumptions we make off the equilibrium path are important because they affect incentives. The standard refinements (Banks 1991, 14–7) rely on finding the type with the largest incentive to deviate from the equilibrium path; players believe that the out-of-equilibrium message (action) is sent by this type. I use this intuition to define out-of-equilibrium beliefs. If nations take aggressive actions (A attacks, B retaliates) out of equilibrium, then they are believed to be the highest type, $\theta^*_m = 1$. I assume that the lowest types take out-of-equilibrium passive actions (A does not attack, B does not retaliate, C does not intervene). These refinements are important for constructing the example, but no propositions rely on these assumptions. Note that C's type is revealed if it fights, so beliefs are well defined if a multilateral war occurs.

**Proof of Proposition 2.** Given any set of beliefs, $\mu_a(\theta, m)$, $\mu_b(\theta, m)$, and $\mu_c(\theta, m)$, domestic reelection makes $C$ more likely to intervene than international factors alone predict: $d\gamma(m)/d\psi \geq 0$.

**Proof:** C is indifferent about intervening if

$$\int_{0}^{1} \int_{0}^{1} \mu_a(\theta, att) \mu_a(\theta, ret) [p(\Theta) - q(\Theta)] d\theta_a d\theta_b - k_c + \Psi(\Phi(E[\theta^*_m, MUWAR]) - \Phi(E[\theta^*_m, BIWAR])) = 0.$$ 

Given C's strategy, $s_c(\theta, m, \mu(\theta), E[\theta^*_m, MUWAR] = \mu(\theta) \mu_c(\theta, m) d\theta$. If C intervenes, then, by assumption, its competence is revealed, $E[\theta^*_m, MUWAR] = \theta^*_m$. Therefore, $E[\theta^*_m, MUWAR] = E[\theta^*_m, BIWAR] > 0$. Since $\mu_c > 0$, and $\Phi$ is increasing in expected competence, then $\Psi(\Phi(E[\theta^*_m, MUWAR]) = \Phi(E[\theta^*_m, BIWAR])) > 0$. Notice that this is true even if we relax the assumption that C's type is revealed when a war occurs. If C's type is not revealed, then $E[\theta^*_m, MUWAR] = \mu_c(\theta, m) d\theta_c > E[\theta^*_m, BIWAR]$.

If domestic politics is unimportant, that is, either $\Psi = 0$ or $\Phi(\mu_c)$ is constant in $E[\theta^*_m, \mu_c]$, then the type $\theta_m$ that satisfies the above equality is higher than the type that satisfies the equality if domestic politics matters. Therefore, domestic reelection makes leaders more hawkish.

**Q.E.D.**

**Proposition 3.** For any $\mu_a(\theta, m)$, if a change in an exogenous parameter increases the probability that C intervenes, then it (weakly) decreases the probability that A attacks: $(da(m)/d\gamma(m)) \leq 0$, where $\alpha$ and $\gamma$ are differentiable.

**Proof:** But suppose not. Suppose a change in an exogenous variable, such as bias or $\Psi$, increases the probability of intervention and increases the probability of initiation: $\gamma'(m)$ > $\gamma(m)$ and $\alpha'(m)$ > $\alpha(m)$.

(1) Consider B's decision to retaliate. $\alpha'(m)$ > $\alpha(m)$ implies that $\theta^*_m(m)$ > $\theta^*_m(m)$. Therefore, if A attacks, then on average B fights a less competent type if it resists. In addition, since some types of C intervene that previously did not, and $p(\Theta) > q(\Theta)$, B's payoff for resisting strictly increases:

$$U_a(\text{ret}|\theta^*_m, m) = \int_{\theta^*_0(m)}^{1} \mu_a(\theta, m) p(\Theta) d\theta_a + (1 - \gamma(m)) q(\Theta) d\theta_a - k_c.$$ 

B's payoff if it surrenders is 0. Therefore, the type that is indifferent about retaliating is lower, that is, $\theta^*_m(m) \leq \theta^*_m(m)$, which implies $B(m) \geq B(m)$.

(2) Consider A's decision. In some cases in which B previously backed down, it now fights back. In addition, in some of these wars C now intervenes, whereas previously it did not. Therefore, A's value from attacking is reduced, as
more often aggression ends in war, and more often these wars are multilateral rather than bilateral. Hence, $U_c(\theta_1, m, \theta_2)$ decreases for all $\theta_2$. But $U_c(\theta_1, m, \theta_2)$ is constant for all changes in $\theta_2(m)$ and $\theta_2(m')$. Therefore, the type that is indifferent about attacking is higher following the exogenous change: $\theta_2^{m'}(m') \geq \theta_2^{m}(m)$, which implies $\alpha(m') \leq \alpha(m)$. This contradicts the initial supposition that $\alpha(m') > \alpha(m)$. Q.E.D.

**Lemma 1.** In equilibrium, each nation's expected utility is a weakly increasing, continuous monotone in its type.

**Proof:** I consider only nation C; the proofs for A and B are analogous.

(i) First show that $U_c(\theta_1, m, \theta_2)$ is weakly increasing in $\theta_2$.

$$U_c(\theta_1, m, \theta_2) = 1 - \alpha(m, \theta_2) + (1 - \alpha(m, \theta_2)) p(\theta_2) d\theta_2.$$

where

$$\Phi(E[\theta_2|m, \theta_3]) + \xi(\theta_2, m, \theta_3),$$

Since $p(\theta_2)$ is increasing in $\theta_2$, and $p(\theta_2) > q(\theta_2)$, $\xi(\theta_2, m, \theta_3)$ is weakly increasing in $\theta_2$. Consider two types, $\theta_2 < \theta_2'$. If type $\theta_2'$ mimics type $\theta_2$, then its payoff is at least as good as the lower types: $U_c(\theta_2', m, \theta_3) \geq U_c(\theta_2, m, \theta_3)$. Therefore, $U_c(\theta_2', m, \theta_3) \geq U_c(\theta_2, m, \theta_3)$.

(ii) Second, show that $U_c(\theta_2, m, \theta_3)$ is a continuous function in $\theta_2$. Consider any two arbitrary types, $\theta_2$ and $\theta_2'$, where $\theta_2 = \theta_2 + \epsilon$. In equilibrium, $U_c(\theta_2', m, \theta_3) \geq U_c(\theta_2, m, \theta_3)$, and $U_c(\theta_2', m, \theta_3) \geq U_c(\theta_2, m, \theta_3)$. As $\epsilon \to 0$, $\xi(\theta_2, m, \theta_3) \to \xi(\theta_2', m, \theta_3)$, and $\xi(\theta_2, m, \theta_3) \to 0$. Therefore, as $\epsilon \to 0$, $U_c(\theta_2', m, \theta_3) \to U_c(\theta_2, m, \theta_3)$. This implies that $U_c(\theta_2, m, \theta_3)$ is continuous.

Q.E.D.

**Lemma 2.** If $s_2(\theta_2, m, \theta_3) = 0$, then for all $\theta_2 < \theta_2'$ either (i) $\alpha(m(\theta_2)) \beta(m(\theta_2)) = 0$, or (ii) $s_2(\theta_2, m, \theta_3) = 0$.

**Proof:** Consider the type $\theta_2'$, which does not intervene given its equilibrium message $m$. Now consider another type, $\theta_2 < \theta_2'$, which sends the equilibrium message $m(\theta_2')$ with positive probability. Suppose that, given this message, war occurs with positive probability, $\alpha(m(\theta_2')) \beta(m(\theta_2')) > 0$. If type $\theta_2'$ mimics type $\theta_2$, then $U_c(\theta_2', m, \theta_3) = U_c(\theta_2, m, \theta_3)$, because neither type intervenes given the message $m(\theta_2')$. $\theta_2 < \theta_2'$.

Having sent its equilibrium message $m(\theta_2')$, suppose that type $\theta_2'$ will intervene if a war occurs. The expected payoff for $\theta_2'$ is

$$U_c(\theta_2', m, \theta_3) = \alpha(m(\theta_2')) \beta(m(\theta_2')) \Phi(E[\theta_2|m, \theta_3]) + (1 - \alpha(m(\theta_2')) \beta(m(\theta_2'))) \Psi \Phi(E[\theta_2|m, \theta_3], BIWAR).$$

Because $p(\Phi)$ is increasing in $\theta_2$, type $\theta_2'$ receives a higher payoff than $\theta_2$ if it sends message $m(\theta_2')$ and intervenes should a war occur, $U_c(\theta_2', m, \theta_3) > U_c(\theta_2, m, \theta_3)$. Since $m(\theta_2')$ is an equilibrium message for type $\theta_2'$, $U_c(\theta_2', m(\theta_2')) = U_c(\theta_2', m(\theta_2'))$. Given that $U_c(\theta_2', m(\theta_2')) = U_c(\theta_2', m(\theta_2'))$, this implies $U_c(\theta_2', m(\theta_2')) > U_c(\theta_2, m(\theta_2'))$, but this is a contradiction of lemma 1.

Q.E.D.

**Lemma 3.** If there exists an equilibrium message $m'$ such that $\alpha(m') = 0$, then either (i) all messages are uninformative, or (ii) there exists an alternative message $m''$, such that $\alpha(m'') > 0$, and $E[\theta_2|m'] > E[\theta_2|m''].$

**Proof:** Consider an alternative message $m''$. There are three cases:

(1) $\alpha(m'') = 0$.

(2) $\alpha(m'') = 1$, $\beta(m'') = 0$, and $\Psi(\theta_2, m'') > 0$.

Consider (1). $U_c(\theta_2, m') = 1 + \Psi(\theta_2, m')$ and $U_c(\theta_2, m'') = 1 + \Psi(\theta_2, m'')$ for all $\theta_2$. Therefore, if both messages are sent in equilibrium, then $E[\theta_2|m'] = E[\theta_2|m'']$ and messages are uninformative.

(2) $U_c(\theta_2, m'') = 1 + \Psi(\theta_2, m'')$, and $U_c(\theta_2, m'') = \Psi(\theta_2, m'')$. Therefore, $E[\theta_2|m''] > E[\theta_2|m']$.

(3) $U_c(\theta_2, m'') = 1 - \alpha(m'') + (1 - \alpha(m'') \beta(m'')) \Psi(\theta_2, m'') + \alpha(m'') \beta(m'') \Psi(\theta_2, m'') - k_2 + \Psi(\theta_2, m'')$.

Since $\alpha(m'') > 0$, and $E[\theta_2|m''] > E[\theta_2|m']$. Q.E.D.

**Lemma 4.** If there exists an equilibrium message $m'$ such that $\beta(m') = 0$, then either (i) all messages are uninformative, or (ii) (nongeneric case) there exists an alternative message message $m''$ such that $\alpha(m'') = 0$, $\beta(m'') > 0$ and $E[\theta_2|m''] < E[\theta_2|m']$ or (iii) there exists an alternative message message, such that $\alpha(m'') > 0$, and $E[\theta_2|m''] > E[\theta_2|m']$.

**Proof:** Consider an alternative message $m''$. There are three cases:

(1) $\beta(m') = 0$.

(2) $\alpha(m'') = 0$, $\beta(m'') > 0$, and $\Psi(\theta_2, m'') > 0$.

Consider (1). Note that $\beta(m') = 0$ implies $\alpha(m'') = 1$. $U_c(\theta_2, m'') = \Psi(\theta_2, m'')$, and $E[\theta_2|m''] = E[\theta_2|m']$ for all $\theta_2$. Therefore, if both messages are sent in equilibrium, then $E[\theta_2|m'] = E[\theta_2|m'']$, and messages are uninformative.

(2) $U_c(\theta_2, m'') = \Psi(\theta_2, m'')$, and $U_c(\theta_2, m'') = 1 + \Psi(\theta_2, m'')$. If both messages are sent in equilibrium, then $E[\theta_2|m''] = E[\theta_2|m'] + \Psi(\theta_2, m'')$. Note that this case is nongeneric.

(3) $U_c(\theta_2, m'') = \Psi(\theta_2, m'') - k_2 + \Psi(\theta_2, m'')$. Q.E.D.

Since $\alpha(m'') > 0$, $E[\theta_2|m''] > E[\theta_2|m']$.
Given $s(\theta, \cdot), E[\theta|m^n] \geq E[\theta|m^n, BIWAR]$. This implies that $E[\theta|m^n, BIWAR] < E[\theta|m^n]$. 

Case b: Suppose that having sent message $m^n$ all types intervene: For all $\theta$, such that $\sigma(\theta, m^n) > 0$, $s(\theta, m^n) = 1$. Let $\theta^n$ be the smallest $\theta$ that sends $m^n$ and intervenes. Since $C$’s expected payoff, $U(\theta, m^n) = 1 - \alpha(m^n) + (1 - \alpha(\theta)(\sigma(m^n))\Psi(\theta|m^n)) + \sigma(m^n)\beta(\theta) - k_\theta + \Psi(\theta|m^n)$. Therefore, $\sigma(\theta, \theta^n) > 0$, and $s(\theta, m^n) = 1$ implies that if $\sigma(\theta, \theta^n) > 0$, then $\theta_\theta \leq \theta^n$. Therefore, $E[\theta|m^n] \leq \theta^n \leq \Psi(\theta|m^n).$ $U(\theta, m^n) \geq U(\theta^n, m^n) \leq U(\theta^n, m^n) \geq U(\theta^n, m^n) = U(\theta^n, m^n) = 0$.

Thus, $1 - \alpha(\theta^n) + \alpha(\theta^n)\beta(\theta^n) - k_\theta < 0 \Rightarrow \theta^n < \theta_\theta$. Yet, since $\theta^n$ intervenes, $\theta^n < \theta_\theta \Rightarrow \Psi(\theta^n) \geq \Psi(\theta|m^n)$. This implies that $E[\theta|m^n] > E[\theta|m^n] > E[\theta|m^n, BIWAR]$. Q.E.D.

**An Informative Equilibrium.** In the informative equilibrium, if $C$ is a type above $\theta^n$, then it sends message $m_1 \in M_1$ and intervenes if a war occurs. If $C$ is a low type, $\theta_\theta < \theta_\theta$, then it sends message $m_2 \in M_2$ and does not intervene if a war occurs. If a multilateral war occurs, then the voters learn $C$’s type. If $C$ sends message $m_1$, then the voters believe that its average competence is $E[\theta|m_1] = (1 + \theta^n_1)/2$. This is greater than the average competence of $C$ if message $m_2$ is sent, $E[\theta|m_2] = \theta^n_2/2$. Yet, since types that send message $m_2$ do not intervene, $C$ pays no domestic political cost for not intervening. Types that send message $m_1$ pay a huge domestic political cost for not intervening if a war occurs because voters believe them to be completely incompetent.

More formally, the following strategy profile $(\sigma, s_1, \cdots, s_1, \theta^n, \cdots)$ and beliefs $(\mu)$ constitute a sequential equilibrium if there exists $\theta^n \in (0, 1)$ that satisfies equations 1 and 2 below:

For all $m_1 \in M_1$ and $m_2 \in M_2$, such that $M_1 = M_1 \varnothing \sigma(\theta_1, m_1) > 0$, if $\theta_1 > \theta_1$, $\sigma(\theta_1, m_1) = 0$ otherwise; $\sigma(\theta_1, m_1) = 1$ if $\theta_1 > \theta_1$. $s_1(\theta_1, m_1) = 0$ otherwise; $s_1(\theta_1, m_1) = 0$ otherwise; and $\theta^n(\theta_1, \cdots, m_1) = 1$ if $\theta_1 > \theta^n(\theta_1, \cdots, m_1) = 0$ otherwise for all $i = a, b, c$, where $\theta^n(m)$ satisfies the definitions in proposition 1, and $\mu$ represents posterior Bayesian beliefs.

\begin{equation}
\theta^n(\theta, m_2) \leq \theta^n(\mu(m_2)); \quad (1)
\end{equation}

\begin{equation}
U(\theta, m_1) = U(\theta^n, m_2); \quad (2)
\end{equation}

**Proof:** The beliefs of the players, given their strategies, were characterized above. The optimal strategies, given these beliefs, are uniquely characterized in proposition 1. In characterizing $\theta^n(m_1)$, I assume the out-of-equilibrium belief refinement that the type which does not intervene having sent message $m_1$ is $\theta_1 = 0$.

Given that $s_1$, are optimal strategies, all that remains is to show that $C$’s message-sending strategy is optimal. If type $\theta_1$ sends message $m_2$ and does not intervene if a war occurs, then its expected payoff is $U(\theta, m_1) = 1 - \alpha(m_1) + f_{\theta_1}(\theta, \mu(\theta)) - \theta_1 + \Psi(\mu^2)$, given equation 1, if $C$ sends message $m_2$, its payoff is constant in $\theta_1$. If $C$ sends message $m_1$ and intervenes, however, then its expected payoff, $U(\theta, m_1) = 1 - \alpha(m_1) + f_{\theta_1}(\theta, \mu(\theta)) - \theta_1 + \Psi(\mu^2) + \Psi(\theta_1)$, is strictly increasing in $\theta_1$. If equations 1 and 2 hold, then $\alpha(m, \theta)$ is utility maximizing, given the strategies of the other players. Q.E.D.

**References**


