DOMESTIC POLITICAL AUDIENCES AND THE ESCALATION OF INTERNATIONAL DISPUTES

JAMES D. FEARON *University of Chicago*

International crises are modeled as a political "war of attrition" in which state leaders choose at each moment whether to attack, back down, or escalate. A leader who backs down suffers audience costs *that increase as the public confrontation proceeds. Equilibrium analysis shows how audience costs enable leaders to learn an adversary's true preferences concerning settlement versus war and thus whether and when attack is rational. The model also generates strong comparative statics results, mainly on the question of which side is most likely to back down. Publicly observable measures of relative military capabilities and relative interests prove to have no direct effect once a crisis begins. Instead, relative audience costs matter: the side with a stronger domestic audience (e.g., a democracy) is always less likely to back down than the side less able to generate audience costs (a nondemocracy). More broadly, the analysis suggests that democracies should be able to signal their intentions to other states more credibly and clearly than authoritarian states can, perhaps ameliorating the security dilemma between democratic states.*

An international crisis occurs when one state resists a threat or demand made by another, with both taking actions that suggest that the dispute might be decided militarily. Crises are frequently characterized as "wars of nerves." Measures such as troop deployments and public threats make crises *public* events in which domestic audiences observe and assess the performance of the leadership. For reasons linked to this public aspect of crises, state leaders often worry about the danger that they or their adversary might become locked into their position and so be unable to back down, make concessions, or otherwise avoid armed conflict.

In this article I model an international crisis as a political "war of attrition." The formalization is motivated by an empirical claim, namely that crises are public events carried out in front of domestic political audiences and that this fact is crucial to understanding why they occur and how they unfold. I characterize crises as political attrition contests with two defining features. First, at each moment a state can choose to attack, back down, or escalate the crisis further. Second, if a state backs down, its leaders suffer *audience costs* that increase as the crisis escalates. These costs arise from the action of domestic audiences concerned with whether the leadership is successful or unsuccessful at foreign policy (Fearon 1990, 1992; Martin 1993).

The formalization has three major benefits. First, it helps answer an important question about the origins of war that is missed in the informal literature and begged by existing formal models of crisis bargaining. Briefly, if fighting entails any cost or risk, then rational leaders would not choose war until they had concluded that attack was justified by a sufficiently low chance of an acceptable diplomatic settlement. Thus another way to ask the question "Why do wars occur?" is to ask what leads states to abandon the hope of a cheaper, nonmilitary resolution. A theoretical answer requires us to explain how a state with rational leaders would *learn*. During a crisis, how do leaders come to revise their beliefs about an opponent so that attack is preferred to holding out for concessions? I shall argue that neither the informal nor the formal literature on international conflict supplies satisfactory answers.

The answer suggested here is that audience costs are an important factor enabling states to learn about an opponent's willingness to use force in a dispute. At a price, audience costs make escalation in a crisis an informative although noisy signal of a state's true intentions. They do so in part by creating the possibility that leaders on one or both sides will become locked into their position and so will be unable to back down due to unfavorable domestic political consequences. I find that in the model, a crisis always has a unique *horizon*—a level of escalation after which neither side will back down because both are certainly locked in, making war inevitable. Before the horizon is reached, the fear of facing an opponent who may become committed to war puts pressure on states to settle. The model thus captures a common informal story about international crises—that their danger and tension arise from the risk of positions hardening to the point that both sides prefer a fight to any negotiated settlement.

The second major benefit of the formal analysis is a set of comparative statics results that provide insights into the dynamics of international disputes. The strongest and most striking of these bear on the question of which state is more likely to concede in a confrontation. I find that regardless of the initial conditions, the state more sensitive to audience costs is always less likely to back down in disputes that become public contests. The intuition is that the greater the domestic cost for escalating and then backing down, the more informative is the signal of escalation and the less escalation is required to con-

vey intentions. A stronger domestic audience thus allows a state to signal its true preferences concerning negotiated versus military settlements more credibly and more clearly.

This result and the audience cost mechanism underlying it suggest hypotheses about how state structure might influence crisis bargaining. For example, if actions such as mobilizing troops create larger audience costs for democratic than for authoritarian leaders, then democratic states should be less inclined to bluff or to try "limited probes" in foreign policy—to make military threats and then back off if resistance is met. More broadly, stronger domestic audiences may make democracies better able to signal intentions and credibly to commit to courses of action in foreign policy than nondemocracies, features than might help ameliorate "the security dilemma" (Herz 1950; Jervis 1978) between democratic states.

The comparative statics results also speak to the question of how relative military capabilities and relative interests influence the outcomes of international disputes. Conventional wisdom suggests that the state with inferior military capabilities, or with fewer "intrinsic interests" at stake, is more likely to back down (e.g., George and Smoke 1974, 556–61; Jervis 1971; Snyder and Diesing 1977, 189–95). Surprisingly, in the model, neither the balance of forces nor the balance of interests has any direct effect on the probability that one side rather than the other will back down once both sides have escalated. The reason is that in choosing initially whether to threaten or to resist a threat, rational leaders will take into account observable indices of relative power and interest in a way that tends to neutralize their impact if a crisis ensues. For example, a militarily weak state will choose to resist the demands of a stronger one only if it happens to be quite resolved on the issues in dispute and so is relatively willing to escalate despite its military inferiority. The argument implies that observable aspects of capabilities and interests should strongly influence who gets what in international politics but that their impact should be seen more in uncontested positions and faits accomplis than in crises. Which side backs down in a crisis should be determined by relative audience costs and by *unobservable*, privately known elements of states' capabilities and resolve.

The third major benefit of the analysis is slightly more technical. The model clarifies how international crises differ structurally from the classical war of attrition studied by economists and theoretical biologists (Maynard Smith 1982, chap. 3; Tirole 1989, chap. 8). In the classical case, two firms (or animals) compete for control of a market (or territory) that is not large enough to support both at a profit. The competition lasts until one or both players "quit." International crises are analyzed here as a war of attrition that differs from the classical model in two important respects. First, in crises state leaders possess an additional option beyond continuing the contest or quitting—*they can always choose to attack*. Second, whereas in the classical war of attrition both sides pay costs for continuing the contest, in international crises it is empirically more plausible to assume that only the side that backs down suffers audience costs.[1] The existence of a military "outside option" along with audience costs proves to have major consequences for strategic behavior. Together they create the possibility of "lock in" and thus give crises a horizon. More technically, whereas the classical war of attrition has an infinity of (asymmetric) equilibria involving delay, the game studied here has a unique equilibrium distribution on outcomes up to the horizon.

First, I briefly review the relevant formal literature and also elaborate the theoretical puzzle: Given incentives to misrepresent, how can states involved in a dispute rationally reach the conclusion that the opponent would prefer war to backing down? I then informally discuss possible answers, arguing for the centrality of domestic audience costs, model an international crisis as a political attrition contest to examine the logic of equilibrium behavior, and, finally, draw some general conclusions.

THE THEORETICAL PUZZLE

The costs and risks of war supply states with strong incentives to locate nonmilitary settlements that both sides would prefer to a fight. Most often, it seems, their efforts are successful: *very* few international disagreements become wars. This may seem unsurprising at first glance. One might expect that given the incentives to avoid war, state leaders who disagree on some issue could simply tell each other what they would be willing to accept rather than fight, and then choose a mutually acceptable bargain. The problem, however, is that states can also have strong incentives to misrepresent their willingness to fight in order to gain a better deal. Given these incentives, quiet diplomatic exchanges may be rendered uninformative about a state's preferences. For example, in the Cuban Missile Crisis Kennedy did not ask Khrushchev what he would do if the United States were to impose a blockade or to attack the missile sites in Cuba: answers would have been almost worthless as indicators, due to Khrushchev's incentives to misrepresent (and Kennedy may also have had an incentive not to tip his hand) (cf. Wagner 1989, 197).

States in a dispute thus face a dilemma. They have strong incentives to learn whether there are agreements both would prefer to the use of force, but their incentives to misrepresent mean that normal forms of diplomatic communication may be worthless. I argue that international crises are a response to this dilemma. States resort to the risky and provocative actions that characterize crises (i.e., mobilization and deployment of troops and public warnings or threats about the use of force) because less-public diplomacy may not allow them credibly to reveal their own preferences concerning international interests or to learn those of other states.

To support this claim it must be shown how such actions can credibly reveal that a state would prefer

using force to making concessions. In particular, how is it that actions like mobilization and public warnings allow learning? If states can have incentives to misrepresent their willingness to use force, why should such actions be taken as credible indicators?

For the most part, the informal literature on international conflict and the causes of war takes it as unproblematic that actions such as mobilization "demonstrate resolve." The literature has focused instead on how psychological biases may impair a leader's ability to interpret crisis signals (e.g., Lebow 1981; Snyder and Diesing 1977, chap. 4; Jervis, Lebow, and Stein 1985). The prior question of how a rationally led state would learn in a crisis, given incentives to misrepresent, has not been answered in a theoretically thorough or satisfactory way.

Consider the inference problem faced by a state whose adversary in a dispute has just mobilized troops. If rational, the state's leaders should increase their belief that the adversary will fight only if a high-resolve adversary is more likely to mobilize than an adversary that in fact prefers backing down to war. Thus, if mobilization is to convey information and allow learning, it must carry with it some cost or disincentive that affects low-resolve more than high-resolve states. In Spence's (1973) terms, mobilization (or any other move in a crisis) must be a *costly signal* if it is to warrant revising beliefs. Costless signals, which often include private diplomatic communication and sometimes more public measures, will be so much "cheap talk," since a state with low resolve may have no disincentive to sending them.[2]

To explain how states learn in a crisis, we need to know what makes escalation or delay costly for a low-resolve state that in fact prefers making concessions to military conflict. It is tempting to answer "the risk of war", but this would beg the question since we are trying to establish how this risk arises in the first place. I shall argue that the role of domestic political audiences has typically been crucial for generating the costs that enable states to learn. First, however, I briefly review how the published formal literature on crisis bargaining has addressed the issue.

A number of studies have developed models in which states rationally update their beliefs about an adversary's resolve in the course of a crisis (Bueno de Mesquita and Lalman 1992; Fearon 1990, 1992; Kilgour 1991; Morrow 1989; Nalebuff 1986; Powell 1990; Wagner 1991). Though this is not always apparent, the mechanism that enables learning in each case is costly signaling.

While updating of beliefs occurs in these models, they actually do not address the question of how and why states might rationally come to conclude that fighting was preferable to holding out for concessions. The reason is that almost all of the models have *finite horizons*: the modeler exogenously determines that one of the states in the game will have a final choice between backing down or fighting. In effect, one player will ultimately have no choice but to "take it or leave it," and this restriction creates a cost for escalation. In actual crises, by contrast, whenever a state has the option of attacking it also has the option of delaying or doing nothing. If there are horizons in actual crises they arise *endogenously*, as a consequence of the fact that for some reason waiting ultimately becomes an undesirable choice. Models that exogenously fix a horizon cannot explain why a state would choose to use force (and thus why wars occur) because they cannot explain what makes force preferable to holding out for concessions by the other side.

There are two partial exceptions to this argument. Nalebuff's (1986) and Powell's (1990) models of nuclear brinkmanship have something like an infinite horizon: they allow states to escalate in a crisis indefinitely, until one side backs down or nuclear war occurs. However, in these models states never *choose* to attack. Instead, war can occur only as a result of an accident beyond either side's control. Thus these formalizations cannot and were never intended to explain why states would *consciously* choose to abandon peace for war.

Historically, war has virtually always followed from the deliberate choices of state leaders, if not always as the result they originally intended (Blainey 1973; Howard 1983). Since this pattern seems likely to continue even in the nuclear age, it makes sense to ask how states could reach the conclusion that attack was worth choosing.

AUDIENCE COSTS, STATE STRUCTURE, AND LEARNING IN INTERNATIONAL DISPUTES

I shall consider several types of costs that could serve to make the actions that states take in crises informative about their actual willingness to fight. I argue that while there are several plausible candidates that may play a role in specific cases, audience costs are probably most important and characteristic of crisis bargaining. I shall then discuss variations in audience costs across regime types, suggesting that they may be most significant in states where foreign policy is conducted by an agent on behalf of a principal, as in democracies.

Signaling Costs in Crises

Two sorts of costs that leaders face for backing down in a crisis should be distinguished. First, there is the domestic and international price for conceding the issues at stake, which is the same regardless of when concessions are made or after how much escalation. Second, there are whatever *additional* costs are generated in the course of the crisis itself. By the costly signaling argument, only such added costs can convey new information about a state's resolve. To ask what enables learning in a crisis—and thus why some states ultimately choose to attack—is to ask what makes escalating and then backing down worse for a leader than simply conceding at the outset.

There are a number of possible mechanisms,

grouped here into three categories: physical costs, costs linked to the risk of accidental or preemptive war, and international and domestic audience costs.

The first class includes the financial and organizational costs of mobilizing and deploying troops and also simple impatience ("time preferences") on the part of state leaders. The economic burden of mobilization is sometimes significant enough to convey information about resolve. The fact that Israel's economy cannot bear full mobilization for very long may make Israeli mobilization unusually informative (Shimshoni 1988, 110). One could also argue that the enormous costs of Desert Shield, given well-known U.S. budget constraints, helped make the deployment a (partially) credible indicator of Bush's preferences (Fearon 1992, 153–54). But since the early nineteenth century, the financial costs of mobilization rarely seem the principal concern of leaders in a crisis, particularly in comparison to how their performance looks to domestic and foreign audiences. In few cases do financial costs seem to be what makes crises into *political* "wars of nerves."[3]

For similar reasons, pure time preferences appear less significant a signaling mechanism in crises than in buyer–seller bargaining and other economic examples.[4] Under time preferences, delay in a crisis would be a costly signal because a leader with a high value for settlement versus war is relatively more impatient to enjoy whatever benefits a resolution would allow. If state leaders are sometimes impatient for a deal, it seems more often due to domestic political pressures (e.g., American elections or Gorbachev or Yeltsin's need for cash) than to a pure preference by the leader for "territory today rather than next month."

The second class of signaling costs concerns risks of war that are generated in some direct way by crisis escalation. Schelling's famous notion of the "threat that leaves something to chance" falls into this category (1960, chap. 8). Schelling suggested that in nuclear crises, at least, escalation or delay might create a risk of war resulting from something other than the deliberate choices of state leaders (e.g., a mechanical mishap or an unauthorized launch). If such risks exist, then escalation in a crisis will be a costly signal of resolve, since the risks are less worth running for a state with low interest in the issues at stake.[5]

In the Cuban Missile Crisis, American decision makers did indeed worry about the risk of war stemming from a mechanical or a command-and-control accident (e.g., Blight and Welch 1990, 109, 311). But even in this most intense of all nuclear crises (a "best case" for the threat-that-leaves-something-to-chance argument), the key decision makers were much more concerned about risks of war connected to what the other side would *choose* to do. While the risk of accidental war may contribute to crisis learning, it rarely, if ever, seems to provide the main cost of escalating a dispute.[6]

First-strike advantages, or incentives for preemptive war, provide a more appealing explanation in this class. If escalating a crisis entails a risk the other side will conclude that concessions are unlikely enough to justify seizing a first-strike advantage, then escalation might be a costly signal of resolve. By running a real risk of preemption, delay in a crisis might credibly reveal a high willingness to fight rather than concede.

This mechanism appears to have figured in some historical cases. For example, part of what made the Russian mobilization in 1914 an informative signal of Russia's willingness to fight was that it was undertaken in the knowledge that it would increase Germany's incentive to choose preemptive war (Fearon 1992, chap. 5).[7] In theory, however, first-strike advantages could also have the opposite effect. A state might conclude that since the adversary has not so far availed itself of a first strike, it must be *more* willing to back down than initially anticipated. Further, major concerns about loss of first-strike advantage do not seem common in case-evidence on international crises, and even when such concerns are present, as in 1914, they often compete with worries about the political disadvantages of going first.[8]

While each of the preceding mechanisms may well foster crisis learning in some cases, I would argue that none fits with our intuitive sense of what it is that makes international crises into political wars of nerves. The reason is that none of these mechanisms recognizes the public aspect of crises, the fact that they are carried out in front of political audiences evaluating the skill and performance of the leadership. In prototypical cases (e.g., the standoff leading to the 1991 Gulf War, the Cuban Missile Crisis, and July 1914), a leader who chooses to back down is (or would be) perceived as having suffered a greater "diplomatic humiliation" the more he had escalated the crisis. Conversely, our intuition is that the more a crisis escalates, the greater the perception of diplomatic triumph for a leader who "stands firm" until the other side backs down.

Political audiences need not and do not always have this pattern of perceptions and reactions: they are social conventions that are at times resolved differently. For example, leaders of small states may be *rewarded* for escalating crises with big states and then backing down, where they would be castigated for simply backing down. Standing up to a "bully" may be praised even if one ultimately retreats.[9] Nonetheless, at least since the eighteenth century leaders and publics have typically understood threats and troop deployments to "engage the national honor," thus exposing leaders to risk of criticism or loss of authority if they are judged to have performed poorly by the relevant audiences. Two illustrations follow, both taken from eighteenth-century diplomacy. While a wealth of similar examples are available from the nineteenth and twentieth centuries, these earlier cases suggest that political audiences have mattered in international confrontations for a long time.

The Seven Years War (1757–64) between France and Britain was preceded by several years of "crisis" diplomacy—threats, warnings, and troop mobilizations and deployments (Higonnet 1968; Smoke 1977, chap. 8). In response to French demands on the Ohio

River Valley, the Duke of Newcastle chose in late 1755 to send two regiments to America to impress the French with British resolve. The decision distressed several of Newcastle's colleagues and ambassadors, who seem to have felt that the action engaged the honor of the king and so committed the cabinet to a warlike course, perhaps unnecessarily. One wrote, "It requires great dexterity to conduct [these diplomatic and military moves] in such a manner to maintain the honor of King and Nation" (Higonnet 1968, 79). In a later interview with Rouillé, the French minister of foreign affairs, the British ambassador reported that the minister "complain'd very much of the licentiousness of our Publick papers in exaggerating things beyond measure which only served to irritate and stir up animosity amongst the lower sort of People in both Nations without a just cause" (p. 80). This complaint suggests that even in nondemocratic, eighteenth-century France, a minister could be concerned with what I have called domestic audience costs: it seems that British pamphlets could have the effect of increasing Rouillé's costs for acceding to British demands (as they increased Newcastle's domestic costs for ceding French demands).

About 35 years later, Britain and Spain nearly went to war over an obscure incident involving alleged Spanish insults to British seamen who had landed on Vancouver Island, along with competing claims on the territory (Manning 1904). Once again, both states resorted to troop mobilizations, forceful diplomatic notes, and public threats. There is strong evidence that these moves created significant domestic audience costs for Prime Minister William Pitt: "With an election imminent, the Opposition was ready to make the most of any of the Government's mistakes in negotiating" (Norris 1955, 572). Pitt's vote for navy credits in Parliament and his government's publication of an account of the Vancouver incident led opposition politicians almost to clamor for appropriate satisfaction of British honor and right. Pitt would have faced serious domestic political costs for backing down, much larger than if he had chosen initially to pursue a less public and aggressive line in the dispute.[10]

The notion that troop movements and public demands or threats "engage the national honor"—thus creating audience costs that leaders would pay if they backed down—continues strongly through the nineteenth and twentieth centuries. Such costs can be classified according to whether the audience that imposes them is domestic or international. Relevant domestic audiences have included kings, rival ministers, opposition politicians, Senate committees, politburos, and, since the mid-nineteenth-century, mass publics informed by mass media in many cases. Relevant international audiences include a state's opponent in the crisis and other states not directly involved, such as allies. Here the costs of escalating and then backing down would be felt indirectly through injury to the state's reputation for threatening the use of force only when serious.

Leaders engaged in disputes appear to worry about both international and domestic audiences. Domestic audience costs may be primary, however. Backing down after making a show of force is often most immediately costly for a leader because it gives *domestic* political opponents an opportunity to deplore the *international* loss of credibility, face, or honor.[11] Because governments are far more likely to be deposed or to lose authority due to internal political developments than due to foreign conquest and because opposition groups frequently condition their activities on the international successes and failures of the leaders in power, domestic audiences may provide the strongest incentives for leaders to guard their states' "international" reputations. Audience costs thus figure in a domestic system of incentives that encourages leaders to have a realist concern with their state's "honor" and reputation before international audiences.

Agency Relations and Audience Costs

As noted, audience costs have a strongly conventional aspect; how they are felt and implemented depends on shared perceptions and expectations in a society. Nevertheless, the historical norm seems to have domestic political audiences punishing a leader who concedes after having deployed troops more than one who concedes outright. Why this norm? My theoretical results will suggest a possible explanation, which I anticipate here in order to develop some broader points about how audience costs vary across types of states.

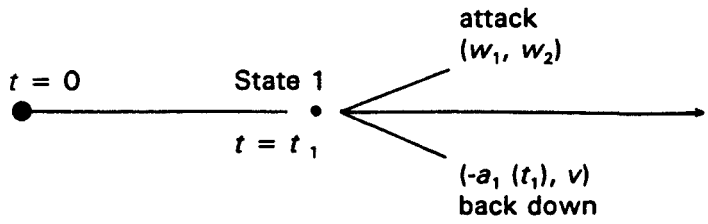
Equilibrium analysis of the crisis model reveals that a state's ex ante expected payoff in a dispute is *increasing* in the degree to which escalatory moves create audience costs for the state's leadership. The reason, in brief, is that greater audience costs improve a state's ability to commit and to signal resolve.

Thus both democratic and nondemocratic leaders should have an incentive to represent that they will pay added domestic political costs for "engaging the national honor" and subsequently backing down. The extent to which such representations are believable, however, depends on the nature of the domestic political institutions that the leadership faces. In democracies, foreign policy is made by an agent on behalf of principals (voters) who have the power to sanction the agent electorally or through the workings of public opinion. By contrast, in authoritarian states the principals often conduct foreign policy themselves. The result here suggests that in the former case, if the principal could design a "wage contract" for the foreign policy agent, the principal would want to commit to punishing the agent for escalating a crisis and then backing down. On the other hand, principals who conduct foreign policy themselves may not be able credibly to commit to self-imposed punishment (such as leaving power) for backing down in a crisis.[12]

Examples of the apparent effect and import of forceful public speeches by democratic and nondemocratic leaders suggest that this argument is at least

FIGURE 1

A Schematic Representation of the Crisis Game



plausible. Repeatedly, leaders in democratic countries have been able credibly to jeopardize their electoral future by making strong public statements during international confrontations. A few prominent examples are Lord Salisbury's speeches during the Fashoda crisis of 1898, Lloyd George's Mansion House speech during the Agadir crisis of 1911, Kennedy's televised speech announcing the presence of Soviet missiles in Cuba, and Bush's many declarations on Kuwait in 1990 (including the "this will not stand" remark).[13] By contrast, even more forceful public bluster by authoritarian leaders (e.g., Hitler, Khrushchev, Mao, and Saddam Hussein) appears to create fewer credible audience costs, and to have correspondingly lower value as signals of intent. For example, it was quite difficult for Western observers to know what to conclude about Saddam Hussein's willingness to fight from his many strong public refusals to pull out of Kuwait in the Fall of 1990. One reason, I would argue, is that it was difficult to know what, if any, added domestic political costs such a tyrant would suffer for making concessions at the last minute.

This is not to suggest that authoritarian states are completely unable to generate audience costs in international confrontations or that democracies can invariably do so. On the one hand, nondemocracies may evolve institutional arrangements in foreign policy that give domestic audiences an ability to sanction decision makers. For example, the politburo after Stalin could sanction the paramount Soviet leader, and eighteenth- and nineteenth-century monarchs could replace unsuccessful ministers. Moreover, since the price of losing power is often greater for a dictator than for an elected leader, a weak or unstable authoritarian regime might be able to create significant *expected* audience costs in a crisis. On the other hand, in democracies the existence of multiple politically relevant audiences may make it difficult for foreign leaders to gauge the costs created by public statements or actions, particularly before elections. The idea that democratic leaders on average have an easier time generating audience costs is advanced here as a plausible working hypothesis that has interesting theoretical and empirical implications.

INTERNATIONAL CRISES AS POLITICAL ATTRITION CONTESTS

I shall describe a model of a crisis as a political attrition contest, informally discuss equilibrium behavior when the states know each other's values for conflict, and then develop the main results on equilibrium with uncertainty about resolve.

Two states—1 and 2—are in dispute over a prize worth $v > 0$. The crisis occurs in continuous time, starting at $t = 0$. For every finite $t \geq 0$ before the crisis ends, each state can choose to attack, quit, or escalate. The crisis ends when one or both states attack or quit. "Escalate" can be thought of either as simply waiting or as taking actions such as mobilizing or preparing troops.

Payoffs are given as follows. If either state attacks before the other quits, both receive their expected utilities for military conflict, w_1 and w_2. These are the states' values for war, incorporating military expectations, values for objects in dispute, and costs for fighting. They can be thought of as levels of "resolve," since the higher w_i, the higher the risk of war state i is willing to run in hope of attaining the prize. Throughout, I shall suppose that neither w_1 nor w_2 is greater than 0.[14]

If state i quits the crisis before the other has quit or attacked, then its opponent j receives the prize while i suffers audience costs equal to $a_i(t)$, a continuous and strictly increasing function of the amount of escalation with $a_i(0) = 0$. I will often consider the linear case $a_i(t) = a_i t$, where $a_i > 0$ is a parameter indicating how rapidly escalation creates audience costs for state i. Figure 1 schematically depicts the structure of the contest, with payoffs indicated in the case that state 1 quits or attacks first at time t_1.[15]

Call this game, with common knowledge of all parameters, G and the subgame beginning at t, $G(t)$. A pure strategy in G is a rule s_i specifying a finite time $t \geq t'$ to quit or attack in every $G(t')$, for all $t' \geq 0$. I shall write $\{t, \text{attack}\}$ for the subgame strategy "escalate up to t, then attack" and $\{t, \text{quit}\}$ for "escalate up to t, then quit."

Equilibrium under Complete Information about Resolve

If the states knew each other's levels of resolve, they could in principle look ahead and anticipate what would happen if they were to escalate and create a crisis. For example, in the linear case where $a_i(t) = a_i t$, they would see that ultimately the audience costs for backing down would be so large that neither would quit: doing so would be strictly worse than attacking. In other words, both states would eventually become "locked in."

However, they would also notice that one side would become locked in *before* the other (excepting certain symmetric cases). Suppose, for example, that audience costs increase linearly and at the same rate on both sides, and that state 1's value for war is higher than state 2's. Then if both sides escalated the crisis, state 1 would reach the point where its leaders preferred conflict to backing down sooner than state 2 would. At this point, state 1 has in effect committed itself not to back down, while state 2 still prefers making concessions to a fight. Thus a rational state 2 would have to back down at this point. Anticipating this at the outset, state 2 would quit immediately rather than pay the larger audience costs that would go with publicly observable escalation.[16]

Thus with complete information about resolve, no public crisis occurs. Instead, if audience costs increase at the same rate on both sides, the state with the lower value for conflict immediately cedes the prize rather than incur costs above those associated with the loss on the issue. If audience costs increase at different rates, then the side with weaker audience effects may be forced to concede even if it has a higher value for war (since it may require more escalation to commit itself to fight).

This equilibrium logic mirrors a logic found in many analytic and diplomatic historical discussions of international disputes. It is often argued that in crises that do not become wars, states look ahead and the side expecting to do worse in military conflict then backs down. But the standard argument does not work by itself: audience costs are required. If it were known that the state with the higher value for war nonetheless preferred making some concessions to a fight, why should the state with the lower value for war necessarily back down? If both prefer concessions to a fight, how can either make a credible threat to go to war and why should this be related to their values for military conflict? Increasing audience costs supply an answer. The state with the higher value for war may be able, in a public crisis, to reach more quickly the point where it prefers conflict to paying the audience costs of backing down.

Incomplete Information about Resolve

With complete information, no public war of nerves occurs because the ultimate outcome can be seen in advance. In a rationalist framework, international crises occur precisely because state leaders *cannot* anticipate the outcome, owing to the fact that adversaries have private information about their willingness to fight over foreign policy interests and the incentive to misrepresent it. I now consider equilibrium in the model in which the states have private information about their willingness to go to war.

Three main theoretical results are developed. Proposition 1 establishes that in any equilibrium in which a crisis may occur, the crisis has a unique *horizon*—an amount of escalation after which neither state will back down and war is certain. Proposition 2 characterizes the set of equilibria in which a crisis may occur. Proposition 3 asserts that in this set, the probability distribution on outcomes is unique up to the horizon time. Throughout, the equilibrium concept is a modification of perfect Bayesian equilibrium (Fudenberg and Tirole 1991) for an infinite game, with an additional restriction ruling out strongly optimistic off-equilibrium-path inferences. Details on the solution concept, along with all proofs, are given in the Appendix.[17]

Preliminaries. Formally, suppose that each state knows its own level of resolve w_i but knows only the distribution of its adversary's resolve w_j. For $i = 1, 2$, let w_i be distributed on the interval $W_i \equiv [\underline{w}_i, 0]$, $\underline{w}_i < 0$, according to a cumulative distribution function F_i that has continuous and strictly positive density f_i.[18] I refer to the crisis game with this information structure and all other elements assumed to be common knowledge as Γ.

In informal terms, f_1 and f_2 represent the states' precrisis beliefs about each other's value for war on the issue in dispute. For example, the more weight f_1 puts on values of w_1 that are close to 0 (as opposed to very negative), the greater is state 2's initial belief that 1 has a relatively high willingness to fight rather than make concessions.

Crises in the Model have a Unique Horizon. This proposition is developed by way of two lemmas, which also help to make clear the logic of strategic choice in the model. I begin with a definition. A crisis has a *horizon* if there is a level of escalation such that neither state is expected to quit after this point is reached. Formally, let $Q_i(t)$ be the probability that state i quits on or before t in some equilibrium. Then $t_h > 0$ is a horizon for Γ if in this equilibrium t_h is the minimum t such that neither $Q_1(t)$ nor $Q_2(t)$ increase for $t > t_h$. If a horizon exists, war has become inevitable by the time it is reached.

Lemma 1 establishes that in any equilibrium in which a crisis may occur, there must exist a horizon. Horizons are thus shown to arise *endogenously*, as a consequence of the equilibrium choices by the states involved. The intuition for this result is straightforward in the case of linearly increasing audience costs: eventually the price of backing down will be so great that even the least-resolved type of state would prefer to attack rather than quit. But the result holds even if "maximum escalation" (arbitrarily large t) will not

create large-enough audience costs to commit every type of state to war.

LEMMA 1. *In any equilibrium of* Γ *in which both states choose to escalate with positive probability, there must exist a finite horizon* $t_h < \infty$.

Lemma 2 characterizes the behavior of states that choose a strategy that could lead to war. The first part establishes that in equilibrium no state will choose to attack *before* a horizon time. When there is no advantage to striking first, a state unwilling to make concessions will want to delay attack as long as there is any chance that the other side will back down, thus avoiding the risk of an unnecessary military conflict while maximizing the chance of a "foreign policy triumph."

The second part of the lemma shows that a state will choose the strategy of escalating to the horizon and then attacking if and only if its privately known level of resolve, w_i, is sufficiently large. Thus crises in the model *separate* states according to their unobservable willingness to fight over the issues: a highly resolved state credibly reveals its motivation by choosing an unyielding crisis-bargaining strategy. This gives it a greater chance of prevailing if the crisis ends peacefully but at the (unavoidable) price of a greater risk of war (cf. Banks 1990). It follows that for outside observers, crisis outcomes *must* be unpredictable to a significant degree. Comparative resolve strongly influences the outcome, but the fact that resolve is privately known (and unobservable) gives rise to public crises in the first place.

LEMMA 2. *In any equilibrium of* Γ *with* t_h *as the horizon and in which escalation may occur, (1) if state* i *chooses* {t, *attack*}, *it must be the case that* $t \geq t_h$; *and (2) state* i *will choose* {t, *attack*} *where* $t \geq t_h$ *if* $w_i > -a_i(t_h)$ *and only if* $w_i \geq -a_i(t_h)$ *(for* $i = 1, 2$).

Lemma 2 implies that the ex ante (precrisis) probability that state i will choose a strategy involving attack is $1 - F_i(-a_i(t_h))$, the prior probability, that $w_i \geq -a_i(t_h)$. Thus we can write state i's ex ante expected utility for escalating up to a horizon time t_h and then backing down as $u_i(t_h) = F_j(-a_j(t_h))v + (1 - F_j(-a_j(t_h)))(-a_i(t_h))$.[19] The function $u_i(\cdot)$ proves to play an important role in defining equilibrium strategies and establishing uniqueness. It is easily shown that $u_i(t)$ is continuous and strictly decreasing and that if audience costs increase "enough" with t there is a unique level of escalation $t_i^* > 0$ such that $u_i(t_i^*) = 0$. Loosely, if the opponent j can generate sufficient audience costs by escalating a crisis, there will be a unique level of escalation such that state i would be indifferent between backing down at $t = 0$ and at t_i^*, were this the horizon. I assume in what follows that the states are able to use escalation to generate audience costs at least this large.[20]

Proposition 1, which follows from lemmas 1 and 2 and from the observations about $u_i(\cdot)$, establishes that if a horizon exists it is unique and is defined as $t^* = \min\{t_1^*, t_2^*\}$. To give a bit of intuition, if a crisis were expected to have a horizon longer than t^*, low-resolve states would prefer to quit immediately rather than "bluff" up to t^*, so t^* could not be the true horizon. On the other hand, if the crisis were expected to have a shorter horizon than t^*, then at least one side would have incentives to bluff by escalation that would make equilibrium unsustainable at t^*: the signal sent by escalating would not be informative enough.

PROPOSITION 1. *Let* t_i^* *be the unique solution to* $u_i(t) = 0$ *for* $i = 1, 2$ *and let* $t^* = \min\{t_1^*, t_2^*\}$. *For any equilibrium of* Γ *in which escalation occurs with positive probability, the horizon must be* t^*.

Equilibrium Strategies and Beliefs. Proposition 2 details the incomplete-information crisis game's equilibria that involve escalation. It indicates that there is a family of substantively identical equilibria: all have t^* as the horizon and have essentially identical behavior in the crisis up to t^*. After t^*, the states may choose any time to attack; the payoff structure leaves this open, not incorporating any incentives for either military delay or an immediate strike. (For ease of exposition, I give normal form strategies that can satisfy the perfection requirements detailed in the Appendix.)

PROPOSITION 2. *Take the labels 1 and 2 such that* $t^* = t_2^* \leq t_1^*$. *Let* $k_1 \equiv u_1(t^*) \geq 0$. *The following describes* equilibrium strategies *for state* $i = 1, 2$ *as a function of type,* w_i:

For $w_i \geq -a_i(t^*)$, *state* i *plays* {t, *attack*} *with any* $t > t^*$.

For $w_i < -a_i(t^*)$, *state* i *plays* {t, *quit*}, *where* t *is chosen according to any pure strategies that yield the cumulative distributions*

$$\mathcal{Q}_1(t) = \frac{1}{F_1(-a_1(t^*))} \frac{a_2(t)}{v + a_2(t)}$$

for state 1, and

$$\mathcal{Q}_2(t) = \frac{1}{F_2(-a_2(t^*))} \frac{k_1 + a_1(t)}{v + a_1(t)}$$

for state 2, both on the interval $[0, t^*]$.

The states' beliefs in equilibrium *are given as follows.*

For $t \leq t^*$, *state* i *believes that the probability* $j \neq i$ *will not back down (i.e.,* $Pr(w_j \geq -a_j(t^*)|t)$) *is*

$$\frac{v + a_i(t)}{v + a_i(t^*)}.$$

For $t > t^*$, *state* i*'s beliefs follow by Bayes' Rule in accord with the opponent's strategy for attacking. For any* $t > t^*$ *off the equilibrium path, let* i *believe that* $w_j > -a_j(t^*)$ *and is distributed according to* F_j, *truncated at* $-a_j(t^*)$.

PROPOSITION 3. *In any equilibrium of* Γ *in which escalation may occur, the equilibrium distribution on outcomes before the horizon time* t^* *implied by proposition 2 is unique.*

An Informal Description of Equilibrium Behavior. Equilibrium behavior in the incomplete information game has the essential features of a war of nerves. At the outset, one side is expected to prefer to make concessions quietly, without a public contest. This state concedes with some probability (k_1/v) at $t = 0$. If it does not make concessions, then its adversary immediately raises its estimate of the state's willingness to fight, and the war of nerves begins. Neither side knows whether or exactly when the other might be locked in by increasing audience costs, but beliefs that the other prefers war to making concessions steadily increase as audience costs accumulate. The reason is that states with low resolve are increasingly likely to have backed down, the more the crisis escalates. Ultimately, in crises that reach the horizon, the only sorts of states remaining have relatively high values for war on the issue. At this point, both sides prefer conflict to backing down, and both know this: attack thus becomes a rational choice.

At a price, then, audience costs enable the states to learn about each other's true willingness to fight over the interests involved in the dispute.[21] The price is paid in two ways. First, a state may escalate or delay for a time and then quit when its adversary matches it. Though the state is still unsure if the adversary really *would* be willing to fight rather than make concessions, its belief that this is possible has increased and it finds it worthwhile to cut its losses. Second, two states may escalate up to the horizon and then fight, even though one or both would have preferred making immediate concessions rather than this outcome. The dilemma created by private information and incentives to misrepresent is that neither can reliably learn that the other would be willing to go this far without taking actions that have the effect of committing both sides to a military settlement.

One further feature of equilibrium in the model deserves comment before I turn to more specific comparative statics results. The more a crisis escalates, the less likely is either side to back down (regardless of precrisis beliefs). In technical terms, the hazard rate is decreasing: the probability that one's opponent will quit after (say) five escalatory moves is less than the probability that the state will quit after four moves. Thus, as escalation proceeds, states in the model gradually become more pessimistic about the likelihood that the adversary will concede after the next round, and outside observers become increasingly concerned that war may be "inevitable."

AUDIENCE COSTS, CAPABILITIES, AND INTERESTS IN INTERNATIONAL DISPUTES

Comparative statics analysis of the equilibrium yields theoretical insights into how three variables affect state behavior and crisis outcomes. I consider in turn the impact of audience costs, relative military capabilities, and relative interests. For expositional convenience, I discuss the case of linearly increasing audience costs, $a_i(t) = a_i t$.

Audience Costs

A striking feature of the equilibrium behavior just described is that the state less able to generate audience costs (lower a_i) is always more likely to back down in disputes that become public contests. This holds regardless of the value of the prize to either side and regardless of the states' initial beliefs about the other's resolve. Thus if actions such as mobilization generate greater audience costs for democratic than for nondemocratic leaders, we should find the democracies backing down significantly less often in crises with authoritarian states.[22]

By itself, intuition can justify the opposite prediction quite easily. One might think that the side less sensitive to its domestic audience would fear escalation less. Knowing this and fearing large costs of retreat, the side with a stronger domestic audience might then be more inclined to back down. But this argument misses the signaling value of escalation for a state with a powerful domestic audience. While such a state may be reluctant to escalate a dispute into a public confrontation, *if it does choose to do so* this is a relatively informative and credible signal of willingness to fight over the issue. That is, the greater the costs created by escalation for a leader, the more likely the leader is to be willing to go to war conditional on having escalated a dispute. Conversely, escalation by a state that will suffer little domestically for backing down says less about the state's actual willingness to fight.[23]

This dynamic has several further implications. First, the signaling and commitment value of a stronger domestic audience helps a state on average, by making potential opponents more likely to shy away from contests and more likely to back down once in them. In the model, a state's ex ante expected payoff increases with its audience-cost rate a_i. This result provides a rationale for why, ex ante, both democratic and authoritarian leaders would want to be able to generate significant audience costs in international contests.

Second, if democratic leaders tend to face more powerful domestic audiences, they will be significantly more reluctant than authoritarians to initiate "limited probes" in foreign policy. Showing this formally requires that we add structure to the model analyzed here, which does not represent an initial choice of one state to challenge or threaten the other. When such an option is added—say, state 1 chooses whether to accept the status quo or to challenge state 2—it is easily shown that the less sensitive state 1 is to audience costs, the greater the equilibrium probability that the state will try a limited probe.[24]

Third, when large audience costs are generated by escalation, fewer escalatory steps are needed credibly to communicate one's preferences. (Formally, the expected level of escalation decreases with a_1 and a_2.) Thus crises between democracies should see signifi-

cantly fewer escalatory steps than crises between authoritarian states—an empirically supported prediction (Russett 1993, 21).

Finally, the equilibrium results bear on the question of how regime type influences the risk that a crisis will escalate to war. When two states in the model have the same audience cost rates $a = a_1 = a_2$, the risk of war conditional on a crisis occurring proves to be *independent* of a, other things being equal (cf. Nalebuff 1986).[25] As audience-cost rates diverge, the high-audience-cost state becomes more likely to escalate, while the lower-audience-cost state becomes more likely to back down. The net effect on the risk of war may be positive or negative, although it is positive for a broad range of plausible parameter values. For example, whenever the distribution of w_1 is uniform, the risk of war (given a crisis) *strictly increases* as a_1 increases above a_2.[26] The model thus suggests a theoretical mechanism that could conceivably help explain the observation that crises between democracies and nondemocracies are more war-prone than are crises between democracies (Chan 1984; Russett 1993). In the model, democratic leaders have a structural incentive to pursue more escalatory, committing strategies when they face authoritarians than when they face fellow democrats, and this can generate a greater overall chance of war.

Relative Capabilities and Interests

Two of the most common informal claims about state behavior in international crises are that (1) the militarily weaker state is more likely to back down and (2) the side with fewer "intrinsic interests" at stake is more likely to back down. These arguments are problematic. If relative capabilities or interests can be assessed by leaders prior to a crisis and if they also determine the outcome, then we should not observe crises between rational opponents: if rational, the weaker or observably less interested state should simply concede the issues without offering public, costly resistance. Crises would occur only when the disadvantaged side irrationally forgets its inferiority before challenging or choosing to resist a challenge (Fearon 1992, chap. 2).

A second striking result from the equilibrium analysis is that observable measures of the balance of capabilities and balance of interests should be unrelated to the relative likelihood that one state or the other backs down in crises where both sides choose to escalate.

In formal terms, observable capabilities and interests influence the distribution of the states' values for going to war and thus the states' initial beliefs about each other's willingness to fight (f_1 and f_2). For example, the more the balance of military power favors state 1, the more state 1—and the less state 2—is initially expected to be willing to use force. Regarding interests, the more the issues in dispute are initially thought to be important for, say, state 1, the more state 1 is initially expected to be willing to fight rather than back down.

In equilibrium, the initial distributions of the states' values for war have a direct influence on the probability that one state or the other will concede without creating a crisis. In accord with intuition, the weaker state 2 is militarily or the less its perceived stake, the more likely it is to cede the prize without offering visible resistance.[27] However, if it *does* choose to escalate, then the odds that state 2 rather than state 1 will back down in the ensuing contest,

$$\frac{a_1 v + a_1 a_2 t^*}{a_2 v + a_1 a_2 t^*},$$

are not directly influenced by relative capabilities or interests. For example, when the states have the same audience cost rates ($a_1 = a_2$), they are equally likely to back down in a crisis and equally likely to go to war, regardless of ex ante indices of relative power or interests.

Less formally, the result suggests that rational states will "select themselves" into crises on the basis of observable measures of relative capabilities and interests and will do so in a way that neutralizes any subsequent impact of these measures. Possessing military strength or a manifestly strong foreign policy interest does deter challenges, in the model. But if a challenge occurs nonetheless, the challenger has signaled that it is more strongly resolved than initially expected and so is no more or less likely to back down for the fact that it is militarily weaker or was initially thought less interested.

CONCLUSION

International crises are a response to a dilemma posed by two facts about international politics: (1) state leaders have private information about their willingness to use force rather than compromise, and (2) they can have incentives to misrepresent this information in order to gain a better deal. In consequence, quiet diplomatic exchanges may be insufficient to allow states to learn what concessions an adversary would in truth be willing to make. I have argued that states resolve this dilemma by "going public"—by taking actions such as troop mobilizations and public threats that focus the attention of relevant political audiences and create costs that leaders would suffer if they backed down. Though there are exceptions, the historical norm seems to have domestic audiences punishing or criticizing leaders more for escalating a confrontation and then backing down than for choosing not to escalate at all.

A game-theoretic analysis showed that such audience costs allow states to learn about each other's willingness to fight in a crisis, despite incentives to misrepresent. When escalation creates audience costs for both sides, states revise upward their prior beliefs that the other is willing to use force as the crisis proceeds. If escalation reaches a certain level (the "horizon"), both states prefer fighting to backing

down, and both know this. At this point, attack becomes a rational choice.[28]

Equilibrium analysis yielded several novel propositions about how audience costs, relative capabilities, and relative interests influence the outcomes of international confrontations. Some broader implications follow.

A substantial literature in international relations argues that international anarchy, combined with states' uncertainty about each other's motivations, is a powerful cause of international conflict (Glaser 1992; Herz 1950; Jervis 1978; Waltz 1959, 1979). Unsure of each other's intentions, states arm and take actions that may make others less secure, leading them to respond in kind. States' inability to commit themselves to nonaggressive policies under anarchy may exacerbate, or even make possible, such "security dilemmas."[29]

The results of my analysis suggest that domestic political structure may powerfully influence a state's ability to signal its intentions and to make credible commitments regarding foreign policy. If democratic leaders can more credibly jeopardize their tenure before domestic audiences than authoritarian leaders, they will be favored in this regard. For example, in the model examined here, high-audience-cost states require less military escalation in disputes to signal their preferences, and are better able to commit themselves to a course of action in a dispute.

This observation provides a theoretical rationale that might help explain why the quality of international relations between democracies seems to differ from that between other sorts of states. If democracies are better able to communicate their intentions and to make international commitments, then the security dilemma may be somewhat moderated between them. For example, the leaders of a democratic state that is growing in power may be better able to commit themselves not to exploit military advantages that they will have in future, so reducing other states' incentives for preventive attack.[30] Likewise, alliance relations between democracies may be less subject to distrust and suspicion if leaders would pay a domestic cost for reneging on the terms of the alliance, so "violating the national honor" in the eyes of domestic critics.[31]

One tradition within realism argues that democratic leaders are at a disadvantage in the game of realpolitik: domestic constraints reduce their freedom to maneuver and so may prevent them from playing the game as hard or as subtly as it may require (e.g., Morgenthau 1956, 512–26). However, as Schelling (1960) observes, in bargaining a player can benefit from having fewer options and less room to maneuver. I have shown how the presence of a politically significant domestic audience can improve a democratic leader's ability to commit to a course of action *and* to signal privately known preferences and intentions in a clear, credible fashion. These are advantages that could help in the game of realpolitik and might also make democracies better able to cope with the security dilemma.

APPENDIX

A formal statement of the solution concept used for the incomplete information game follows. Because Γ has a continuum of information sets, standard definitions of perfect Bayesian equilibrium (Fudenberg and Tirole 1991) and sequential equilibrium (Kreps and Wilson 1982) do not apply. I propose an adaptation of perfect Bayesian equilibrium, adding a refinement criterion that rules out some optimistic interpretations of out-of-equilibrium play. To avoid measure-theoretic complications, attention is restricted to pure strategy equilibria. Throughout, $i = 1, 2$ and $j \neq i$.

I begin by defining a Bayesian Nash equilibrium for the normal form version of Γ. Here, a pure strategy for state i is a map s_i: $W_i \rightarrow \mathbb{R}^+ \times \{\text{quit, attack}\}$, where $\mathbb{R}^+$ is the set of nonnegative reals. Using F_i, every s_i induces a unique pair of cumulative distributions $Q_i(t)$ and $A_i(t)$, which are the probabilities that state i quits or attacks by t if i follows s_i. By the properties of cumulative distribution functions, $Q_i(t)$ and $A_i(t)$ are increasing, right-continuous, and have well-defined left-hand limits for all t (Billingsley 1986, 189). Let

$$Q_i^-(t) \equiv \lim_{s \rightarrow t^-} Q_i(s).$$

"State w_i's" expected payoff for $\{t, \text{quit}\}$, given s_j, is then

$$U_i^q(t, w_i) \equiv Q_j^-(t)v + (Q_j(t) - Q_j^-(t))((v - a_i(t))/2) + A_j(t)w_i + (1 - Q_j(t) - A_j(t))(-a_i(t))$$

or, if $Q_j(t)$ is nonatomic at t,

$$U_i^q(t, w_i) = Q_j(t)v + A_j(t)w_i + (1 - Q_j(t) - A_j(t))(-a_i(t)).$$

Similarly, type w_i's expected payoff for $\{t, \text{attack}\}$ given s_j is

$$U_i^a(t, w_i) \equiv Q_j^-(t)v + A_j(t)w_i + (1 - Q_j^-(t) - A_j(t))w_i = Q_j^-(t)v + (1 - Q_j^-(t))w_i.$$

DEFINITION. *$\{t', quit\}$ ($\{t', attack\}$) is a* best reply for type w_i given s_j *if*

$t' \in argmax_t\, U_i^q(t, w_i)$ *and* $U_i^q(t', w_i) \geq max_t\, U_i^a(t, w_i)$

($t' \in argmax_t\, U_i^a(t, w_i)$ *and* $U_i^a(t', w_i) \geq max_t\, U_i^q(t, w_i)$).

DEFINITION. *(s_1, s_2) is a* Bayesian Nash equilibrium *for the normal form version of Γ if (1) $\{F_i, s_i\} \Rightarrow \{Q_i(t), A_i(t)\}$, and (2) under s_i, every type w_i chooses a best reply, given s_j.*

Just as the normal form version of the complete information game G has multiple Nash equilibria, so are there multiple Bayesian Nash equilibria for Γ. However, many of these require states to choose strategies that do not seem optimal or sensible in the dynamic (extensive form) setting. These are ruled out by the "perfection" requirements I shall give.

In the extensive form, a complete pure strategy in Γ is a map s_i: $\mathbb{R}^+ \times W_i \to \mathbb{R}^+ \times \{\text{quit, attack}\}$, with the restriction that if $s_i(t', w_i) = \{t, \text{quit}\}$ or $\{t, \text{attack}\}$, then $t' \leq t$. For all $t' \geq 0$, define the "continuation game" $\Gamma(t')$ as follows: (1) payoffs are as in Γ, except beginning at t'; and (2) "initial beliefs" are given by a cumulative distribution function $F_i(\cdot; t')$ on W_i. A strategy s_i implies a strategy for state i in every continuation game $\Gamma(t')$; call this $s_i|t'$. Further, using $F_i(\cdot; t')$, $s_i|t'$ induces a pair of unique "conditional" cumulative probability distributions $Q_i(t|t')$ and $A_i(t|t')$, analogous to $Q_i(t)$ and $A_i(t)$ already defined. From these, expected payoff functions for $\Gamma(t')$, $U_i^q(t|t', w_i)$ and $U_i^a(t|t', w_i)$ follow as before.

We can now define a weak extensive form solution concept requiring that (1) s_1 and s_2 induce Bayesian Nash equilibria in every continuation game $\Gamma(t)$, and (2) beliefs $F_i(\cdot; t)$ are formed whenever possible using Bayes' Rule and s_i, while $F_i(\cdot; t)$ can be anything when Bayes' Rule does not apply.

DEFINITION. $\{(s_1, s_2), F_1(\cdot;\cdot), F_2(\cdot;\cdot)\}$ *is a* perfect Bayesian equilibrium *for* Γ *if*

(A) (s_1, s_2) *induces a Bayesian Nash equilibrium in* Γ *and for all* $t \geq 0$, $(s_1|t, s_2|t)$ *induces a Bayesian Nash equilibrium in* $\Gamma(t)$, *using* $F_1(\cdot; t)$ *and* $F_2(\cdot; t)$; *and*

(B) for all t *such that* t *is reached with positive probability under* s_i *(i.e.,* $Q_i(t) + A_i(t) < 1$*),* $F_i(\cdot; t)$ *is* $F_i(\cdot)$ *updated using Bayes' Rule and* s_i.

This solution concept is weak in the sense that it imposes no restrictions on how states would interpret completely unexpected behavior by the adversary. For instance, if state 1 escalates unexpectedly at time t, the concept allows state 2 to conclude that state 1 is without doubt the *least resolved* type w_1. Further, it would allow state 2 to maintain this belief even as state 1 continued to escalate. Seemingly implausible "optimistic beliefs" of this sort can be used to support a continuum of perfect Bayesian equilibria in Γ for most initial parameter values (with t^* as the maximum possible horizon). The following criterion rules out such optimistic off-equilibrium-path inferences and so refines the set of equilibria.[32] It is stronger than is needed for the proofs that follow, but it has the advantage of a very simple definition:

(C) For all $t > 0$ *such that* $Q_i(t) + A_i(t) = 1$, $F_i(-a_i(t); t) = 0$.

This says that if state i escalates beyond t when it was expected to have quit or attacked prior to time t, then state j believes that i's value for war w_i is at least as great as i's value for backing down at time t. In the text and in what follows, I refer to a pair of strategies (s_1, s_2) and a system of updated beliefs $F_i(\cdot; t)$ that satisfy A, B, and C as an *equilibrium* of Γ. I now proceed to proofs of lemmas and propositions in the text, starting with several observations (proofs for observations 2 and 4 are straightforward and are omitted).

OBSERVATION 1. *Suppose that in an equilibrium of* Γ, $Q_j(t)$ *is atomic at* t'. *Then* $\{t', \text{quit}\}$ *and* $\{t', \text{attack}\}$ *are never best replies for state* i *for any* w_i *and are chosen with zero probability in equilibrium.*

Proof. Suppose to the contrary that in some equilibrium type w_i chooses $\{t', \text{quit}\}$ where $Q_j(t') > Q_j^-(t')$. State w_i then receives an ex ante expected payoff of

$$Q_j^-(t')v + (Q_j(t') - Q_j^-(t'))((v - a_i(t'))/2) + A_j(t')w_i + (1 - Q_j(t') - A_j(t'))(-a_i(t')).$$

By right continuity of $Q_j(t)$ and $A_j(t)$, the deviation $\{t' + \varepsilon, \text{quit}\}$, $\varepsilon > 0$, yields an expected payoff arbitrarily close to

$$Q_j(t')v + A_j(t')w_i + (1 - Q_j(t') - A_j(t'))(-a_i(t'))$$

as ε approaches 0, which is strictly greater than the payoff for $\{t', \text{quit}\}$. Thus $\{t', \text{quit}\}$ cannot be a best reply for any type w_i. An identical argument applies for $\{t', \text{attack}\}$. Q.E.D.

Observation 1 implies that if in some equilibrium t_h is the horizon, it cannot be that both states choose to quit with positive probability at t_h. Further, we can now write state w_i's equilibrium ex ante expected payoff for $\{t, \text{quit}\}$ as

$$U_i^q(t, w_i) = Q_j(t)v + A_j(t)w_i + (1 - Q_j(t) - A_j(t))(-a_i(t))$$

and state w_i's equilibrium ex ante expected payoff for $\{t, \text{attack}\}$ as $U_i^a(t, w_i) = Q_j(t)v + (1 - Q_j(t))w_i$.

OBSERVATION 2. $U_i^a(t, w_i)$ *increases with* $Q_j(t)$ *for all* w_i. *Thus in any equilibrium of* Γ *no type of state* i *will choose* $\{t, \text{attack}\}$ *(and* $A_i(t) = 0$*) whenever there exists a* $t' > t$ *such that* $Q_j(t') > Q_j(t)$.

OBSERVATION 3. *Suppose* $t_h > 0$ *is a horizon in some equilibrium of* Γ *in which escalation may occur. Then for all* $\varepsilon > 0$ *state* i *quits with positive probability in the interval* $[t_h - \varepsilon, t_h)$ *for* $i = 1, 2$.

Proof. If t_h is a horizon, then by definition at least one state (say, i) must quit with positive probability in the interval $[t_h - \varepsilon, t_h]$ for all $\varepsilon > 0$. I first show that this implies that the same must hold for j. If the contrary is true, then in some equilibrium, there must exist a $t' < t_h$ such that for all $t \geq t'$, $Q_j(t) = Q_j(t')$. By observation 2, $A_j(t) = 0$ for $t < t_h$, so $U_i^q(t, \cdot) = Q_j(t)v + (1 - Q_j(t))(-a_i(t))$ for $t < t_h$. $U_i^q(t, \cdot)$ is strictly decreasing in t whenever $Q_j(t)$ is constant and less than 1, so if the contrary is true and $Q_j(t') < 1$, then no type of i would be willing to choose $\{t, \text{quit}\}$ for any $t > t'$, contradicting the hypothesis that t_h is the horizon. If $Q_j(t' < t_h) = 1$, then $\{t < t', \text{quit}\}$ is not a best reply for any w_i, implying that $Q_i(t') = 0$. It follows that t' must equal 0—$\{t, \text{quit}\}$ with $0 < t \leq t'$ never being a best reply for any w_j—so escalation does not occur with positive probability, contradicting the hypothesis. Thus both states must quit with positive probability in the interval $[t_h - \varepsilon, t_h]$ for all $\varepsilon > 0$.

By observation 1, there can be no equilibrium in which both states choose $\{t_h, \text{quit}\}$ with positive

probability. Thus in any equilibrium with t_h as the horizon, at least one state (say, i) quits with positive probability in the interval $[t_h - \varepsilon, t_h)$ for all $\varepsilon > 0$. If observation 3 is false, then it must be possible to have an equilibrium in which $Q_j(t)$ is atomic at t_h but j does not quit with positive probability in an interval $[t_h - \delta, t_h)$ for small-enough $\delta > 0$. But then $Q_j(t)$ will be constant and less than 1 for $t \in [t_h - \delta, t_h)$, so by the same logic as in the last paragraph, i will not be willing to choose $\{t, \text{quit}\}$ with $t \in [t_h - \delta, t_h)$, contradicting the hypothesis. Q.E.D

OBSERVATION 4. *Suppose that* t_h *is the horizon in some equilibrium of* Γ. *By observations 2 and 3, for* $j = 1, 2$, $A_j(t) = 0$ *for* $t < t_h$. *Thus* $\{t > t_h, \text{attack}\}$ *yields an ex ante expected payoff of* $U_i^a(t, w_i) = Q_j(t_h)v + (1 - Q_j(t_h))w_i$, *while* $\{t < t_h, \text{quit}\}$ *yields* $U_i^q(t, \cdot) = Q_j(t)v + (1 - Q_j(t))(-a_j(t))$. *Since* $U_i^q(t, \cdot)$ *is independent of* w_i *for all* t *such that* $\{t, \text{quit}\}$ *is a best reply for state* w_i, $U_i^q(t, \cdot)$ *must equal a constant (call it* k_i).

Proof of Lemma 1. Suppose to the contrary that there exists an equilibrium of Γ in which escalation may occur and in which $Q_i(t)$ is strictly increasing for all t for some state i. By observation 2, $A_j(t) = 0$ for all $t \geq 0$ (since for all $t \geq 0$ there exists a $t' > t$ such that $Q_i(t') > Q_i(t)$). And, by observation 4, i's equilibrium expected payoff for $\{t, \text{quit}\}$ is

$$U_i^q(t, \cdot) = Q_j(t)v + (1 - Q_j(t))(-a_i(t)) = k_i,$$

implying that

$$Q_j(t) = \frac{k_i + a_i(t)}{v + a_i(t)} \quad (*).$$

However, because $A_j(t) = 0$ for all t, it must be that

$$\lim_{t \to \infty} Q_j(t) = 1.$$

From (*), this will be possible only if

$$\lim_{t \to \infty} a_i(t) = \infty$$

or if $k_i = v$, both of which generate contradictions. If

$$\lim_{t \to \infty} a_i(t) = \infty,$$

then no type of i will be willing to choose $\{t, \text{quit}\}$ for arbitrarily large t, if $\Gamma(t)$ actually occurred. If $k_i = v$, then $Q_j(0) = 1$, implying that j does not escalate with positive probability. Q.E.D

Proof of Lemma 2. Part 1 follows immediately from observations 2 and 3. As for the second part, fix an equilibrium in which escalation may occur and there exists a horizon t_h. Let T_i be the set of times such that for all $t \in T_i$, $Q_i(t)$ is either atomic or strictly increasing. Observations 3 and 4 imply that i's ex ante expected payoff for $\{t \in T_i, \text{quit}\}$ is

$$\begin{aligned} U_i^q(t, \cdot) &= Q_j(t)v + (1 - Q_j(t))(-a_i(t)) \\ &= Q_j^-(t_h)v + (1 - Q_j^-(t_h))(-a_i(t_h)) = k_i. \end{aligned}$$

Also from observation 4, type w_i's ex ante expected payoff for $\{t > t_h, \text{attack}\}$ is

$$U_i^a(t, w_i) = Q_j(t_h)v + (1 - Q_j(t_h))w_i,$$

which is at least as great as $U_i^a(t_h, w_i)$.

There are now two cases to consider. First, if $Q_j^-(t_h) = Q_j(t_h)$, then i's ex ante expected payoff for $\{t < t_h, \text{quit}\}$, $t \in T_i$, is $Q_j(t_h)v + (1 - Q_j(t_h))(-a_i(t_h))$, which implies that i does better to choose $\{t \geq t_h, \text{attack}\}$ if $w_i > -a_i(t_h)$ and only if $w_i \geq -a_i(t_h)$. Second, if $Q_j^-(t_h) < Q_j(t_h)$, then there exists a $\hat{w}_i < -a_i(t_h)$ such that type $\hat{w}_i$ is indifferent (ex ante) between $\{t > t_h, \text{attack}\}$ and $\{t < t_h, \text{quit}\}$ and thus a measurable set of types $W_i' \equiv (\hat{w}_i, -a_i(t_h))$ that strictly prefer $\{t > t_h, \text{attack}\}$ to $\{t < t_h, \text{quit}\}$. But this is impossible. The action $\{t > t_h, \text{attack}\}$ yields $Q_j(t_h)v + (1 - Q_j(t_h))w_i$, while $\{t_h + \varepsilon, \text{quit}\}$ yields $Q_j(t_h)v + A_j(t_h + \varepsilon)w_i + (1 - Q_j(t_h) - A_j(t_h + \varepsilon))(-a_i(t_h + \varepsilon))$. If $A_j(t_h) \neq 1 - Q_j(t_h)$, then for small enough $\varepsilon > 0$, the quit strategy does strictly better for all $w_i \in W_i'$. If $A_j(t_h) = 1 - Q_j(t_h)$, then all $t > t_h$ are off the equilibrium path. Condition C implies that for $t > t_h$, $F_j(-a_j(t); t) = 0$, so $Q_j(t|t_h) = 0$ in all $\Gamma(t)$ for $t > t_h$. But if j will not quit after t_h, then $\{t > t_h, \text{attack}\}$ cannot be a best reply in the continuation games $\Gamma(t > t_h)$ for types in W_i'. Thus $Q_j^-(t_h) < Q_j(t_h)$ is impossible in any equilibrium with $t_h > 0$ as the horizon, and the first case must hold. Q.E.D.

Proof of Proposition 1. From lemma 2, it follows that in any equilibrium with horizon $t_h > 0$, the ex ante probability that state j chooses $\{t \geq t_h, \text{attack}\}$ is $1 - F_j(-a_j(t_h))$. Thus, using observation 3, $U_i^q(t_h - \varepsilon, \cdot)$ can be made arbitrarily close to $F_j(-a_j(t_h))v + (1 - F_j(-a_j(t_h))(-a_i(t_h))$, which, by consequence, must equal k_i.

Choose labels such that $t_2^* \leq t_1^*$. I show first that t_h cannot be strictly greater than t_2^* in any equilibrium. If it were, then state 2's payoff for $\{t_h - \varepsilon, \text{quit}\}$ would be $k_2 = F_1(-a_1(t_h))v + (1 - F_1(-a_1(t_h)))(-a_2(t_h)) < 0$, which is impossible. Because $A_1(0) = 0$, state 2 can assure itself at least 0 by the strategy $\{0, \text{quit}\}$, and so state 2 would not be willing to choose $\{t, \text{quit}\}$ for any $t > 0$, contradicting observation 3.

Nor can t_h be strictly less than t_2^*. If it were, then both states must expect an equilibrium payoff $k_i = F_j(-a_j(t_h))v + (1 - F_j(-a_j(t_h)))(-a_i(t_h)) > 0$ for $\{t < t_h, \text{quit}\}$. Since for $t \in T_i$, $t < t_h$, $k_i = Q_j(t)v + (1 - Q_j(t))(-a_i(t))$, $k_i > 0$ implies that for both states there must exist a $t_j' \geq 0$ such that $Q_j(t)$ is atomic at t_j' and such that $Q_j(t) = 0$ for all $t < t_j'$. If this were not the case, then for one state, $Q_j(t)$ would require types of j to play $\{t, \text{quit}\}$ when this yielded a payoff of 0 or less, which could not be a best reply for any type. Moreover, it must be the case that $t_1' = t_2'$; if not, then for state i with $t_i' < t_j'$, $\{t_i', \text{quit}\}$ yields a payoff less than or equal to zero. But this contradicts observation 1, since in no equilibrium can both states quit with positive probability at the same time. Thus t_h must equal t_2^* in any equilibrium of Γ and thus, $t^* > 0$ is unique. Q.E.D

Proof of Proposition 2 (Sketch). That the proposed strategies form a Bayesian Nash equilibrium in Γ follows immediately from lemma 2 and the fact that $\mathcal{Q}_j(t)$ is chosen so that all types $w_i < -a_i(t^*)$ are indifferent among $\{t, \text{quit}\}$ for all $t < t^*$ ($k_1 = u_1(t^*)$ and $k_2 = 0$). For the continuation games $\Gamma(t')$, $t' \geq 0$, Bayes' Rule implies that if $\{F_i, s_i\} \Rightarrow \{Q_i(t), A_i(t)\}$, then

$$Q_i(t|t') = \frac{Q_i(t) - Q_i(t')}{1 - Q_i(t') - A_i(t')}$$

and

$$A_i(t|t') = \frac{A_i(t) - A_i(t')}{1 - Q_i(t') - A_i(t')},$$

for $t > t'$ and t' such that $Q_i(t') + A_i(t') < 1$. Notice that when they are defined, $Q_i(t|t')$ and $A_i(t|t')$ are linear transformations of $Q_i(t)$ and $A_i(t)$, respectively. This fact can be used to show that if $\{t, \text{quit}\}$ is a best reply for type w_j given s_i, then it remains so in all continuation games $\Gamma(t' \leq t)$, provided that $Q_i(t') + A_i(t') < 1$ under s_i (and likewise for $\{t', \text{attack}\}$). Thus the strategies given for types given in the proposition form Bayesian Nash equilibria in every continuation game $\Gamma(t')$ up to the earliest time t'' such that for one of the two states, $Q_i(t'') + A_i(t'') = 1$ (if t'' exists). It is straightforward to show that beliefs off the equilibrium path ($t > t''$, if t'' exists) accord with condition C.

Proof of Proposition 3 (Sketch). By proposition 1, the horizon must be t^* in any equilibrium in which escalation may occur. The only question, then, is whether there are equilibrium distributions on outcomes up to t^* that differ from $\mathcal{Q}_1(t)$ and $\mathcal{Q}_2(t)$. Arguments similar to those for observations 1 and 3 establish that any equilibrium quit distributions must be nonatomic and strictly increasing on $[0, t^*)$. But $\mathcal{Q}_1(t)$ and $\mathcal{Q}_2(t)$ are the only nonatomic strictly increasing distributions that make types $w_i < -a_i(t^*)$ willing to quit at any $t \in [0, t^*)$ and also support t^* as the horizon.

Notes

Presented at the annual meetings of the American Political Science Association, Washington, 1993. Thanks to Atsushi Ishida, Andy Kydd, Robert Powell, Jim Morrow, Matthew Rabin, and Barry Weingast for comments.

1. Studying the "diplomacy of insults," Barry O'Neill (n.d.) independently developed an attrition model of international contests that focuses on this same second feature.

2. For the original discussion of costly signaling in economics, see Spence 1973. On cheap talk (which may be informative in some contexts), see Farrell 1988; Crawford and Sobel 1982; Rabin 1990. The crisis signals discussed herein are atypical in that they create costs that are paid only if the signaler takes a certain future action ("backing down") rather than regardless of what the signaler does in the future (as in Spence's classical case). One implication is that these signals can have a commitment (or "bridge-burning") effect. For a discussion of costly signaling in crises, see Fearon 1992, chaps. 3 and 4, and for the seminal treatment of signaling in international relations, see Jervis 1970.

3. For example, the financial costs of sustained mobilization do not appear as a significant factor in the case studies found in Betts 1987; George and Smoke 1974; Lebow 1981; or Snyder and Diesing 1977.

4. Time preferences or opportunity costs are the main rationale given for the costs of delay in models of buyer–seller bargaining. For an overview of these models, see Fudenberg and Tirole 1991, chap. 10. On time preferences in crisis bargaining, see also Morrow 1989, 949.

5. Schelling's own account views the threat that leaves something to chance primarily as a tactical move available to both sides, rather than as a mechanism for revealing private information about resolve (see Fearon 1992, chap. 3; Maxwell 1968).

6. On the importance of distinguishing between "loss of control" in a crisis due to pure accident and "loss of control" due to unanticipated but deliberate decisions, see Powell 1985.

7. Trachtenberg (1991) shows convincingly that the German need to mobilize and attack before Russian mobilization was far advanced was known in both Berlin and St. Petersburg.

8. See the citations in n. 3. I consider how first-strike advantages affect crisis bargaining and escalation in work-in-progress.

9. A number of examples from Balkan conflicts are discussed in Fearon 1992, 184–85.

10. According to Norris, "Pitt was conscious that he must negotiate an agreement acceptable to the new parliament when it met in the autumn, or face political annihilation" (1955, 574). The terms that would have been acceptable to the parliament were (intentionally) made harsher by Pitt's public escalation of the crisis. Spain ultimately backed down and Pitt was much praised for his diplomatic triumph.

11. Significant American examples include the heat Acheson took for "losing China" and Johnson's and Nixon's fears about domestic criticism for sending the wrong signal to the communists over Vietnam) (see, e.g., Gelb and Betts 1979, 220–26.

12. Alternatively, domestic audiences may draw harsh inferences about a leader's competence if the leader backs down in a crisis after escalating. If they do, then this would also create an audience cost that would be felt more strongly in democratic states. On the use of incentive schemes to improve an agent's bargaining power, see Katz 1991.

13. For discussion and citations, see Fearon 1992, chap. 3.

14. Nothing important changes if w_1 and w_2 are allowed to be greater than zero but less than the value of the prize. Also, there is no loss of generality in setting both sides' value for the prize equal to v.

15. Payoffs have been defined except for simultaneous quits or attacks, which do not play much of a role in the sequel. If one state chooses to attack at t and the other chooses to quit or attack at the same time, they receive (w_1, w_2). If both quit at time t, state i receives $(v - a_i(t))/2$.

The assumption that the winner gets v independent of the amount of escalation makes the analysis more tractable without discarding a key feature of crises that distinguishes them from the classical war of attrition, namely, that only one side pays the costs of escalation if a player quits. The assumption that w_i does not depend on t is also made for tractability; it would be both interesting and desirable to relax it. Discounting is omitted for simplicity and so that I can focus on the independent impact of audience costs.

16. For all values of w_1 and w_2, there is one other outcome obtainable in a subgame perfect equilibrium, this one involving the play of weakly dominated strategies: both sides choose {0, attack}. If each expects the other to attack immediately, neither has an incentive to deviate. This equilibrium disappears in an alternating-move version of the game. If the two states would be locked in at the same time, then there are three equilibrium outcomes (beyond the weakly dominated one): either one of the two states quits immediately or both play a mixed strategy up to the lock in time.

17. I have omitted proofs of the comparative statics results. These are available on request, along with less-compressed versions of the proofs given here.

18. I assume also that f_1 and f_2 are independent, which is most naturally interpreted to mean that uncertainty is about the opponent's cost–benefit ratio for war rather than about military capability. For a discussion of this issue, see Fearon 1993.

19. This interpretation of $u_i(t_h)$ as i's expected utility for $\{t_h$, quit$\}$ is valid only if j neither quits nor attacks with positive probability at t_h, but the proofs do not depend on the interpretation.

20. If they cannot (e.g., if $a_1(t) = a_2(t) = 0$ for all t), then there may not exist an equilibrium in which learning occurs.

21. In the linear case, as the audience-cost rates a_1 and a_2 approach zero, the horizon time t^* approaches infinity, meaning that an arbitrarily large amount of delay or escalation is required to credibly signal willingness to fight.

22. In formal terms, the probability that state 1 will back down prior to the horizon time is $a_2t^*/(v + a_2t^*)$. The probability that state 2 will do the same, conditional on the crisis occurring (i.e., lasting longer than $t = 0$), is $a_1t^*/(v + a_1t^*)$. Thus if $a_1 > a_2$, state 2 is more likely to back down than state 1, and vice versa. This result holds for *any* precrisis beliefs, f_1 and f_2.

23. The probability that state i will fight conditional on a crisis occurring is $v/(v + a_jt^*)$, $j \neq i$, and t^* proves to rise as a_i falls, implying the result in the text.

24. In the Cold War period the Soviet Union appeared generally more willing than the United States to threaten the use of military force and then back off or moderate on meeting resistance. This, at any rate, is a reading consistent with standard interpretations of the set of major Cold War crises (e.g., Betts 1987; George and Smoke 1974; and Snyder and Diesing 1977). Certainly the United States has used force on many occasions in Latin America and elsewhere, but military probes to gauge other parties' willingness to resist appear uncommon. Maoz and Russett (1992, 253) report that 62% of 271 post-1945 crises between democracies and "autocracies" were initiated by the autocracy—a number that would extremely unlikely to occur if democracies were just as likely to try military probes.

25. Relevant other things will not be equal if democratic leaders tend to have higher (audience) costs for fighting wars than do nondemocratic regimes. Indeed, the same argument that suggests that democratic leaders will suffer more politically for backing down after escalating a crisis suggests that they will be more sensitive to potential war costs. In the model, crises are less likely to escalate to war the greater are the states' costs for fighting (or, equivalently, the lower v).

26. When the distribution of w_1 is logistic up to $w_1 = 0$, the probability of war given a crisis (Pr(war|crisis)) increases as a_1 increases above a_2 whenever the median value of w_1 is sufficiently low. For example, let $F_1(z) = (1 + \exp(-z - m))^{-1}$ for $z < 0$, and $F_1(0) = 1$. If $v = 1$, then the result holds whenever $m > 1$, which means that the typical state 1 is not willing to run 50% risk of war for the prize. Ultimately, for highly asymmetric situations (very large a_1, very small a_2), Pr(war|crisis) begins to decrease with a_1 in the logistic case. For instance, if $v = 1$ and $m = 5$, Pr(war|crisis) is .045 when $a_1 = a_2$, and reaches a maximum at .11 when a_1 is about 50 times greater than a_2.

27. The probability that state 2 backs down at $t = 0$ is k_1/v. A shift in the balance of power (understood as the probability that 1 would win a war) shifts f_1 to the right and f_2 to the left; this has the consequence of increasing k_1. An increase in the intensity of state 1's interests at stake shifts f_1 to the right (without affecting f_2), which also has the consequence of increasing k_1.

28. An important limitation of the attrition model of crises (a limitation common to most other models of "crisis bargaining") is that it gives states only two ways to resolve a dispute peacefully: one side or the other must "back down." While some evidence suggests that many crises in fact have this aspect (Snyder and Diesing 1977, 248), we would like to know why. A natural next step is to consider models with continuous-offer bargaining (e.g. Fearon 1993; Powell 1993).

29. There is in fact a large set of different arguments lumped under the "security dilemma" heading, but this criticism cannot be pursued here. For a critique of standard "security dilemma" reasoning, see Kydd 1993.

30. Schweller (1992) provides some evidence suggesting that democracies neither engage in nor are the targets of preventive war. Fearon (1993) shows that between rationally led states, preventive war arises from the rising power's inability to commit not to exploit the future bargaining advantage it will have.

31. A similar argument about alliances is developed by Gaubatz (1992), who presents evidence indicating that alliances between democracies last longer than alliances involving nondemocracies. See also Fearon 1992, 355.

32. The refinement is in the spirit of Cho and Kreps (1987) D1 criterion. Even without the refinement, comparative statics results are only marginally weakened for the set of perfect Baysian equilibria of Γ.

References

Banks, Jeffrey. 1990. "Equilibrium Behavior in Crisis Bargaining Games." *American Journal of Political Science* 34:579–614.

Betts, Richard. 1987. *Nuclear Blackmail and Nuclear Balance*. Washington: Brookings Institution.

Billingsley, Patrick. 1986. *Probability and Measure*. 2d ed. New York: Wiley & Sons.

Blainey, Geoffrey. 1973. *The Causes of War*. New York: Free Press.

Blight, James, and David Welch. 1990. *On the Brink: Americans and Soviets Reexamine the Cuban Missile Crisis*. New York: Noonday.

Bueno de Mesquita, Bruce, and David Lalman. 1992. *War and Reason*. New Haven: Yale University Press.

Chan, Steve. 1984. "Mirror, Mirror on the Wall, . . . Are Freer Countries More Pacific?" *Journal of Conflict Resolution* 28:617–48.

Cho, In-Koo, and David Kreps. 1987. "Signaling Games and Stable Equilibria." *Quarterly Journal of Economics* 102:179–222.

Crawford, Vince, and Joel Sobel. 1982. "Strategic Information Transmission." *Econometrica* 50:1431–52.

Farrell, Joseph. 1988. "Communication, Coordination, and Nash Equilibrium." *Economic Letters* 27:209–14.

Fearon, James. 1990. "Deterrence and the Spiral Model: The Role of Costly Signals in Crisis Bargaining." Presented at the annual meeting of the American Political Science Association, San Francisco.

Fearon, James. 1992. "Threats to Use Force: The Role of Costly Signals in International Crises." Ph.D. diss., University of California, Berkeley.

Fearon, James. 1993. "Rationalist Explanations for War." Presented at the 1993 annual meeting of the American Political Science Association, Washington.

Fudenberg, Drew, and Jean Tirole. 1991. *Game Theory*. Cambridge: Massachusetts Institute of Technology Press.

Gaubautz, Kurt Taylor. 1992. "Democratic States and Commitment in International Relations." Presented at the annual meeting of the American Political Science Association, Chicago.

Gelb, Leslie, and Richard Betts. 1979. *The Irony of Vietnam: The System Worked*. Washington: Brookings Institution.

George, Alexander L., and Richard Smoke. 1974. *Deterrence in American Foreign Policy*. New York: Columbia University Press.

Glaser, Charles. 1992. "Political Consequences of Military Strategy: Expanding and Refining the Spiral and Deterrence Models." *World Politics* 44:497–538.

Herz, John. 1950. "Idealist Internationalism and the Security Dilemma." *World Politics* 2:157–74.

Higonnet, P. L. R. 1968. "The Origins of the Seven Years' War." *Journal of Modern History* 40:57–90.

Howard, Michael. 1983. *The Causes of Wars*. Cambridge: Harvard University Press.

Jervis, Robert. 1970. *The Logic of Images in International Relations*. Princeton: Princeton University Press.
Jervis, Robert. 1971. "Bargaining and Bargaining Tactics." In *Nomos: Coercion*, ed. J. Roland Pennock and J. Chapman. Chicago: Aldine.
Jervis, Robert. 1978. "Cooperation under the Security Dilemma." *World Politics* 30:167–214.
Jervis, Robert, Richard N. Lebow, and Janice Gross Stein. 1985. *Psychology and Deterrence*. Baltimore: Johns Hopkins University Press.
Katz, Michael. 1991. "Game Playing Agents: Unobservable Contracts as Precommitments." *Rand Journal of Economics* 22:307–28.
Kilgour, D. Marc. 1991. "Domestic Political Structure and War Behavior: A Game-theoretic Approach." *Journal of Conflict Resolution* 35:266–84.
Kreps, David, and Robert Wilson. 1982. "Sequential Equilibrium." *Econometrica* 50:863–94.
Kydd, Andrew. 1993. "The Security Dilemma, Game Theory, and World War I." Presented at the 1993 annual meeting of the American Political Science Association, Washington.
Lebow, Richard Ned. 1981. *Between Peace and War*. Baltimore: Johns Hopkins University Press.
Manning, William Ray. 1904. "The Nootka Sound Controversy." *Annual Report of the American Historical Association*. Washington: GPO.
Maoz, Zeev, and Bruce Russett. 1992. "Alliance, Contiguity, Wealth, and Political Stability: Is the Lack of Conflict among Democracies a Statistical Artifact?" *International Interactions* 17:245–67.
Martin, Lisa. 1993. "Credibility, Costs, and Institutions: Cooperation on Economic Sanctions." *World Politics* 45:406–32.
Maxwell, Stephen. 1968. *Rationality in Deterrence*. Adelphi Paper No. 50. London: Institute for Strategic Studies.
Maynard Smith, John. 1982. *Evolution and the Theory of Games*. Cambridge: Cambridge University Press.
Morgenthau, Hans. 1956. *Politics among Nations*, 2d ed. New York: Knopf.
Morrow, James D. 1989. "Capabilities, Uncertainty, and Resolve: A Limited Information Model of Crisis Bargaining." *American Journal of Political Science* 33:941–72.
Nalebuff, Barry. 1986. "Brinkmanship and Nuclear Deterrence: The Neutrality of Escalation." *Conflict Management and Peace Science* 9:19–30.
Norris, J. M. 1955. "The Policy of the British Cabinet in the Nootka Crisis." *English Historical Review* 70:562–80.
O'Neill, Barry. N.d. "The Diplomacy of Insults." In *Signals, Symbols, and War*. Forthcoming.
Powell, Robert. 1985. "The Theoretical Foundations of Strategic Nuclear Deterrence." *Political Science Quarterly* 100:75–96.
Powell, Robert. 1990. *Nuclear Deterrence Theory: The Problem of Credibility*. Cambridge: Cambridge University Press.
Powell, Robert. 1993. "Bargaining in the Shadow of Power." University of California at Berkeley. Typescript.
Rabin, Matthew. 1990. "Communication between Rational Agents." *Journal of Economic Theory* 51:144–70.
Russett, Bruce. 1993. *Grasping the Democratic Peace: Principles for a Post–Cold War World*. Princeton: Princeton University Press.
Schelling, Thomas. 1960. *The Strategy of Conflict*. Cambridge: Harvard University Press.
Schweller, Randolph. 1992. "Domestic Structure and Preventive War: Are Democracies More Pacific?" *World Politics* 44:235–69.
Shimshoni, Jonathan. 1988. *Israel and Conventional Deterrence: Border Warfare from 1953 to 1970*. Ithaca: Cornell University Press.
Smoke, Richard. 1977. *War: Controlling Escalation*. Cambridge: Harvard University Press.
Snyder, Glenn, and Paul Diesing. 1977. *Conflict among Nations*. Princeton: Princeton University Press.
Spence, A. Michael. 1973. "Job Market Signalling." *Quarterly Journal of Economics* 87:355–74.
Tirole, Jean. 1989. *The Theory of Industrial Organization*. Cambridge: Massachusetts Institute of Technology Press.
Trachtenberg, Marc. 1991. *History and Strategy*. Princeton: Princeton University Press.
Wagner, R. Harrison. 1989. "Uncertainty, Rational Learning, and Bargaining in the Cuban Missile Crisis." In *Models of Strategic Choice in Politics*, ed. Peter Ordeshook. Ann Arbor: University of Michigan Press.
Wagner, R. Harrison. 1991. "Nuclear Deterrence, Counterforce Strategies, and the Incentive to Strike First." *American Political Science Review* 85:727–49.
Waltz, Kenneth. 1959. *Man, the State, and War*. New York: Columbia University Press.
Waltz, Kenneth. 1979. *Theory of International Politics*. Reading, MA: Addison-Wesley.

James D. Fearon is Assistant Professor of Political Science, University of Chicago, Chicago, IL 60637.