doi:10.1017/S0003055410000158

Diplomatic Calculus in Anarchy: How Communication Matters

ROBERT F. TRAGER *University of California at Los Angeles*

When states come to believe that other states are hostile to their interests, they often reorient their foreign policies by realigning alliance commitments, building arms, striking first, mobilizing troops, or adopting policies to drain the resources of states that menace them. This article presents a crisis bargaining model that allows threatened states a wider array of responses than the choice to back down or not. Two implications are that (1) "cheap talk" diplomatic statements by adversaries can affect perceptions of intentions, and (2) war can occur because resolved states decline to communicate their intentions, even though they could, and even though doing so would avoid a war. The model relates the content and quality of diplomatic signals to the context of prior beliefs about intentions and strategic options. In simulations, this form of diplomatic communication reduces the likelihood of conflict.

In the fall of 1876, Russia queried Germany as to what its position would be if Russia were to go to war with Austria-Hungary. Germany replied, as delicately as it possibly could, that it could not guarantee its neutrality in such an eventuality, and that "a lasting weakening of Austria would be contrary to [German] interests."[1] The German chancellor, Otto von Bismarck, thought that this created a "new situation" in the German-Russian relationship.[2] He meant that Germany had increased Russian expectations about Germany's willingness to go to war on behalf of Austria, and perhaps decreased Russia's appraisal of the extent to which Germany would support Russian aims generally.

According to most theories of information transmission between states, this should have been impossible. Germany communicated its position privately through diplomatic channels. Thus, German elites did not stake their reputation before a domestic audience, nor did their actions carry explicit costs, two commonly recognized mechanisms of information transmission.[3] Further, Germany's noncommittal statements can hardly be considered to have so engaged its reputation before the Russian and interstate audiences that anyone could have believed Germany significantly less willing to back down for fear of the repercussions of having been caught in a bluff. Theories of information transmission that rely on the staking of bargaining reputations, therefore, also fail to explain the case.[4] Yet Bismarck's appraisal was correct: thereafter, Russia reckoned on German support for Austria-Hungary.[5]

In this article, I describe a bargaining model in which states have a richer array of options than the previous literature has allowed. This results in a mechanism of information transmission that provides a convincing explanation for this case, and, I shall argue, represents an essential aspect of interstate diplomacy. The anarchic strategic context can provide less resolved states with a disincentive to misrepresenting their intentions. Very often, when states negotiate over important security concerns, there is a danger that a threat (or other noncooperative behavior) will result in a *breach* in relations, not merely with respect to the issue at hand, but also with respect to other aspects of the relationship. In such cases, states can send "costless" signals about their intentions. Information is conveyed by threats because states understand the dangers of altering other states' *perceptions* of their intentions, and yet choose to threaten anyway when they are sufficiently resolved. Alternatively, if foreign policy choices are responsive to perceptions of other states' intentions at all, then explicit threats from one state to another, whether in public or private, can convey information. Thus, this article contributes to the literature on the conditions under which costless diplomatic signals are informative.[6]

A second result of the model is a path by which incomplete information can lead to conflict that is different from the mechanism in other models in the literature. War can occur in some equilibria because resolved states decline to communicate their resolve

Robert F. Trager is Assistant Professor, Department of Political Science, University of California at Los Angeles, 4289 Bunche Hall, Box 951472, Los Angeles, CA 90095-1472 (rtrager@ucla.edu).

I am grateful to Michael Chwe, Maria Fanis, Erik Gartzke, Mike Horowitz, Robert Jervis, Andy Kydd, Helen Milner, Barry O'Neill, Kris Ramsay, Sebastian Rosato, Ken Schultz, Art Stein, Dessie Zagorcheva and seminar participants at the Olin Institute for useful comments and suggestions, as well as to Bob and Judith Terry and Oriel College, Oxford, for a conducive environment in which to write. I bear responsibility for remaining errors and omissions.

[1] Schweinitz (1927, 360). The Russians did not explicitly threaten Austria, but it was at this time that the saying, "the road to Constantinople leads through Vienna" (Rupp 1941, 232) gained popularity in Russian diplomatic and military circles and many Russians and Germans thought that the tense, ongoing Austro-Russian negotiations could lead to war. See Ignatyev (1931, 391), Rupp (1941, 297), Saburov (1929, 82), Schweinitz (1927, 359), and *Grosse Politik*, II, 54–66, 74–9.

[2] Bismarck 1915, 286.

[3] See, for instance, Fearon (1994) and Kydd (1997).

[4] See, for instance, Sartori (2005). See Fearon (1995) for an analysis of the efficacy of costless diplomatic signals.

[5] As the German Ambassador wrote, "nothing was more natural" (Schweinitz 1927, 350–1) than for the Tsar to expect a guarantee of German neutrality, and he remained openly disturbed by the German response for days (Rupp 1941, 204; Schweinitz 1927, 360–364). According to the Russian diplomat Kartsov, the idea of declaring war on Austria "was abandoned" because "Prince Bismarck forewarned that Austria was necessary for Germany for reasons of political balance of power and that Germany would therefore not permit delivering Austria a death-blow" (Rupp 1941, 297). See Trager (2007, chap. 3).

[6] Other works on this subject, in addition to those cited above, include Guisinger and Smith (2002), Jervis (1970), Kurizaki (2004), Kydd (2003), Ramsay (2004), and Schelling (1966; 1980).

even though they could, and even though this would prevent the war. This can occur when one state is highly resolved, and believes that another is also, but believes that the other state does not believe that the first is highly resolved. In such cases, if the first state were to signal its resolve, there would be a high probability that the other state would prepare for conflict. To avoid this likelihood, the first state declines to convey its resolve to the other. Sometimes, however, the second state would have been willing to comply with the first state's demand in order to avoid war. Thus, in this situation, resolved states mimic the behavior of unresolved states in order to catch the other side unprepared. This contrasts with other signaling models in which it is generally unresolved states that would like to imitate the signals sent by resolved states.

This second point is similar to results derived from other recent models that have begun to expand the range of state options beyond those conceived of in traditional bargaining models. Ritter (2004, chap. 2) argues that states sometimes make their alliances secret, in spite of the drawbacks from the point of view of deterrence, to prevent potential adversaries from taking countermeasures. Slantchev (n.d.) presents an ultimatum game in which rejected offers lead to conflicts in which each side must choose a level of effort in the fighting. Because the optimal effort of states depends on each side's perception of the strength of the other side, states sometimes have an incentive to make offers in the preconflict stage that hide their true strength. The model below reveals a similar dynamic in the context of costless diplomatic encounters across a wide range of seemingly disparate contexts, from alliance politics to nuclear brinkmanship.

The article has four sections. In the first, I locate the argument in the literature on communication in international relations. In the second, I argue that the strategic options available to threatened states go well beyond the binary choice to comply or not in traditional models of coercion. The third section presents a model that allows states that are threatened to prepare for conflicts and to choose themselves to engage in conflict. When these preparations are effective, informative costless signaling occurs in plausible equilibria. This section discusses the conditions under which signaling can be effective, and under which wars occur that available but unused signals could prevent. A final section discusses implications of the analysis.

COMMUNICATION, CONSTRUCTION, AND DETERRENCE

Social aspects of the human world exist in language, through which they are created, altered, and reified. Within the social realm, some complex constructs, for instance, those that would be identified as sources of cultural difference, are the products of many agents interacting over long periods of time and consequently change only very gradually in general. But other aspects of the social world, such as the "situation" described by Bismarck, change more quickly, and such change is more often intentional. To bring about this sort of change, attempts at communication must succeed in conveying information. That is, they must be believed, at least partially. This article will not address how symbols acquire their meanings or how individuals are initiated into language groups.[7] Rather, I assume for analytical purposes that a shared and unchanging language exists and ask this second question, namely, how it is that attempts at communication by adversaries can convey information.

In particular, I look at one of the most adversarial aspects of relations between states, when one is attempting to coerce another through a threat to use force.[8] Here, more than elsewhere, we expect communication to be difficult because the state making the threat has an incentive to misrepresent its preferences in order to make the threat appear more credible. In such situations, as Jervis (1970) recognized, analysts distinguish potential sources of information that can be manipulated from those that cannot. The latter can be believed, whereas the former only convey information in certain contexts.

Within the category of information sources that can be manipulated, some messages are credible, even between adversaries, because senders would not *want* to misrepresent their preferences although they could. When the message causes the sender, in some contingencies, to incur costs *as a direct result of sending the message*, such messages are called "costly signals." In game-theoretic terms, the decision to send the message or not has a *direct* effect on the payoff function of the sender. Perhaps the clearest example from international relations is the decision to build arms. This affects the balance of capabilities directly, but if the arms are costly enough,[9] it may also send a signal to an adversary about the state's *intentions*.[10] The adversary may conclude that a less resolved state would be unwilling to pay those costs.[11]

[7] The literature that addresses these questions is too large to cite even a representative sample. Among philosophers, the work of Ludwig Wittgenstein is the most widely appreciated. Within international relations, see for instance, Finnemore and Sikkink (1998), Jervis (1970, 139–224), O'Neill (1999), and Wendt (1999, 313–69).

[8] See Schelling (1966, 70–1) for the definition of coercion as subsuming "deterrence" and "compellence." But see also Morgan (2003, 3) for a discussion of the intimate relationship of these two from an analytical point of view.

[9] If spending on arms does not reach a threshold related to the possible values states could place on the issue in question, it will generally be be uninformative (at least as a *costly* signal).

[10] For a definition of intentions that is similar to the game-theoretic concept of a strategy, see Jervis (1976, 48–57). For another definition of intention, see Bratman (1999).

[11] It is not argued here that states must draw this conclusion when they observe other states choosing to build arms, but merely that such a dynamic is plausible and easily grasped. There is a large literature that discusses costly signaling, especially as it relates to crisis bargaining. See, for instance, Fearon (1995; 1997), Jervis (1970, 28–9), Kydd (1997), Morrow (1989), and Powell (1990). It also may be worth noting that, as Sartori (2005, 58) points out, although "audience costs" have been modeled as costly signals, this is probably best thought of as a modeling shorthand (when voters are not modeled explicitly) for what is really a costless process. See also Ramsay (2004) and Smith (1998).

But messages need not be associated with direct costs for there to be disincentives to lie, and thus for the message to enable to the receiver to learn something from it. Whether such disincentives exist depends on the strategic context. Suppose, for instance, that an employer wants to give a high-paying position to an experienced applicant and a low-paying position to an inexperienced one. If there is no way to verify whether a candidate is telling the truth, the candidate may have a hard time convincing the employer that he or she deserves the high-paying position. On the other hand, if the high-paying position is for a pilot and the low-paying one for a flight steward, the candidate's statements about his or her hours of flight experience may well be at least partially informative. Misrepresenting qualifications in this case may result in his or her flying planes, which for the inexperienced may result in a crash that has negative consequences for both employer and employee. In this employment example, therefore, the two sides have interests in common; communication of private information may be possible. In the game-theoretic literature, such signals are called costless because the actions available to agents and the payoffs associated with material outcomes are all unchanged no matter which message is sent.[12]

Models in the economics literature show that the maximal degree of precision of credible costless messages increases as the interests of sender and receiver grow more aligned.[13] In two-player games, when relationships are zero-sum, no "cheap talk" communication is possible at all. Understanding the degree to which interests are aligned, however, turns out to be far more difficult than it appears at first glance.[14]

Many scholars of international relations have concluded that relations between states are too adversarial for costless communication to occur. For instance, Schelling noted that "words are cheap, not inherently credible when they emanate from an adversary. . . ."[15] Fearon, careful not to overstate the case, argued that "it remains unclear whether cheap talk is important in international disputes. . . ." His work as a whole suggests, however, that, "costless signals. . . have no effect" on international outcomes.[16] Many other scholars have followed their lead. I show below that the preferences of state adversaries are in fact sometimes sufficiently aligned to allow costless communication to occur.

This distinction between costly and costless signals may seem merely technical, with little substantive importance. This is not the case. On it hinges the question of whether diplomacy, particularly that practiced by leaders behind closed doors, can play a role in determining the course of events in a rationalist framework. Macro-level theories of the international system have generally claimed to explain international events without reference to diplomatic communication.[17] This orientation has received support from micro-level theorizing of conflict processes, which has suggested (if not stated outright) that only directly costly activities such as mobilizing troops, building arms, and creating "backdown costs" can convey information. As a result, international relations scholarship has neglected the study of the role of private diplomatic communications, both theoretically and empirically.

There are important exceptions to this characterization of the literature, however, and recent scholarship has shown a renewed interest in the mechanisms of private diplomacy. Sartori (2002; 2005) has demonstrated that private diplomatic communications can be made credible by the desire of states to maintain a bargaining reputation, which makes them hesitant to send misleading signals for fear of being caught in a bluff.[18] Guisinger and Smith (2002) combine two strains of the literature to show that in the presence of a reputational mechanism along the lines of Sartori's, democratic selection of leaders results in an endogenous additional disincentive against bluffing. Kurizaki (2004) shows that if a publicly threatened state will lose face by backing down, then it is sometimes optimal for states to make private threats. Such threats do not increase the probability that the threatened state assigns to the threatener following through, but neither do these private threats cause that probability to go to zero.[19] Kydd (2003) presents a cheap-talk model of third-party mediation.[20]

[12] The term "costly signal" is sometimes used differently from its definition in the game-theoretic literature to refer to communication mechanisms that rely on any disincentives associated with sending the signal, whether direct or arising out of the strategic interaction of agents. Because any signal of resolve that conveys information must result from a disincentive to unresolved types to send the signal, all informative signals must be "costly" in this sense. The staking of private reputations, which convey information because of the "cost" of not being believed in the future (Sartori 2005), and every other mechanism of information transmission are costly signals on this understanding, but not according to the way these terms are understood in the game-theoretic literature. Recent work in political science has used the game-theoretic terminology. In the international relations literature, see Kydd (2003), Ramsay (2004), and Sartori (2005, 51). In other literature see, for instance, Gilligan and Krehbiel (1987).

[13] See Crawford and Sobel (1982), Farrell and Gibbons (1989), and for a nontechnical discussion, Farrell and Rabin (1996). To better understand the distinction between costly and costless signals, it may be useful to consider the following. If sending the signal costs the sender $100 and the receiver of the signal infers something from the sender's willingness to pay $100, this constitutes a costly signal. Suppose the signal costs nothing to send, however, but conveys information that causes the receiver to *take* $100 from the sender. The signal may only have affected the receiver's beliefs because of the anticipated reaction of the receiver and the sender's willingness to suffer the receiver's reaction. Nevertheless, this scenario represents a *costless* signal because the disincentive to send a misleading signal arises from the strategic interaction of the two agents rather than from a direct cost of taking the action itself. This is so even though the ultimate result—the loss or not of $100—is the same.

[14] See Axelrod (1970).

[15] Schelling (1980, 150). Elsewhere, Schelling argues that words used to *frame* an issue affect perceptions of resolve. See Schelling (1966, 35–91).

[16] Fearon (1997, 69).

[17] For instance, Mearsheimer (2001); Schweller (1998); Walt (1987); Waltz (1979).

[18] Jervis (1970, 78–83), Schelling (1966; 1980), and others have also argued in favor of a reputational signaling mechanism in private diplomacy.

[19] See also Fearon (1997, 84).

[20] Ramsay (2004) describes a cheap-talk model of public diplomacy.

Earlier work describes several other mechanisms of private diplomacy.[21] Schelling (1966, 1980) argued that threats sometimes risk an undesired event that neither side directly controls, which causes the threats to convey information. The essence of brinkmanship is that when one climber who is roped to another moves closer to the edge, he or she may slip. By demonstrating a willingness to slip, the climber conveys information about his or her resolve in the issue being negotiated.[22] Schelling's analysis focuses on how engaging in limited forms of conflict can constitute an implicit threat because it demonstrates a willingness to risk even more costly conflict. Private threats may have a similar effect, however, in that they delay resolution of the issue and create a crisis atmosphere in which conflict may be more likely. Schelling's work bears important similarities to the analysis to follow. Here, however, the emphasis will be on the intentional action of actors rather than the partly exogenous danger of sliding off a cliff and on the implications of including the option to prepare in models of the conflict process. The analysis to follow also makes it clear that some of the dynamics of engaging in limited conflict are equally germane to costless signals and thus to private diplomacy.[23]

Of these mechanisms for private information transmission, the staking of bargaining reputations has received the most attention. Despite some scholarship that argues to the contrary, I believe the diplomatic–historical record supports the proposition that the maintenance of a reputation for being true to one's commitments has been an important force in international affairs.[24] As in the case of Germany's tacit threat to Russia in 1876, mentioned above, however, important information is often conveyed by costless diplomacy that cannot be explained by reputational dynamics. In this case and in many others, in making a threat, policy-makers on both sides were not primarily concerned with bargaining reputation. Rather, Germany worried about consequences that would follow *if, as it expected, its private threat were found credible*. Further, in many cases where reputations may appear to have been staked, the mechanism I propose here represents an alternative and sometimes more plausible explanation of communication.

The model in this article contributes to this growing literature by providing a new mechanism of information transmission between states that applies to certain systemic contexts and appeals neither to the costliness of the signal, nor to bargaining reputation. The theory applies equally to messages conveyed in public and in private. It also demonstrates that a similar signaling mechanism is at work in seemingly diverse strategic contexts, such as nuclear brinkmanship and alliance bargaining.

THE STRATEGIC CONTEXT OF ANARCHY

In prominent models in the international relations literature, the strategic options of the target of a threat are limited: it can stand firm or concede. This conceptualization does not correspond to the options and incentives of states in the anarchic international context. If a state comes to believe it cannot achieve its key strategic aims through its current relationship with another state, it may choose to alter that relationship. In particular, rather than merely deciding whether or not to back down, threatened states must decide how to prepare for conflict if they believe a breach with a threatening state is imminent. In such cases, states often reorient their security policies in order to drain resources from the threatening state, and they also tend to form new alliances that are contrary to the threatening state's security interests. In addition, when the target of a threat believes future conflict is more likely, it will sometimes choose to increase arms production, mobilize troops, or strike first. These decisions are often made before the threatening state chooses to back down or follow through on its threat and have consequences whether or not the states involved ultimately go to war.[25]

To illustrate the strategic choices available to Targets, consider the responses of U.S. President Kennedy and Soviet Premier Khrushchev to each other's threats in their June 1961 meeting. Put simply, each threatened to escalate to war if the other did not accept a settlement of the Berlin question favorable to his side. Following the meeting, Khrushchev responded by adopting a new set of policies designed to drain the resources of the United States. On August 1, he "approved most of a KGB plan to create 'a situation in various areas of the world that would favor the dispersion of attention and resources by the United States and their satellites, and would tie them down during the settlement of a German peace treaty and West Berlin.'"[26] For his part,

[21] Jervis (1970) proposes three additional mechanisms through which private threats could convey information. First, if leaders are reticent to lie for moral reasons, credibility may attach to their statements. Second, if a country has a stake in the current functioning of the international system, a reticence to lie may derive from a desire to ensure that states do not too often deceive each other because a baseline of honest communication may be required to maintain the overall systemic equilibrium. Third, lies may result in unwanted "changes in the international environment." Sometimes, if a statement is believed, other actors may act in such a way that the actor making the statement has an additional incentive to follow through on the statement. For instance, if one state professes hostility toward another, the reaction of the second state may make it in the first state's interest to take hostile actions it had not originally planned on taking.

[22] Powell (1988) formalizes this idea in the context of inadvertent war.

[23] The constructivist literature has emphasized the role of behavior in creating norms and shaping expectations, interests, and identities, but has not focused on the communication of intention among adversaries. See, however, Der Derian (1987).

[24] For criticisms of a reputational signaling mechanism, see Jervis (1984), Mercer (1996), Morrow (1994), and Press (2005) .

[25] Works that argue that states react to perceptions of intentions in ways described here include Jervis (1976, chap. 3), Schelling (1966), Schultz (2001), Schweller (1994), and Walt (1987). One work that argues against state responsiveness to the intentions of other states is Mearsheimer (2001). Waltz (1979) does not argue against a causal role for perceptions of intentions. Rather, he argues only for a separate and independent effect of the distribution of capabilities. See Waltz (2003, 53).

[26] Fursenko and Naftali (1999, 138).

hard as it is to imagine today, Kennedy began to take the idea of a nuclear first strike more seriously. He told the Joint Chiefs of Staff that "Berlin developments may confront us with a situation where we may desire to *take the initiative* in the escalation of conflict from the local to the general [nuclear] war level."[27] Thus, because both leaders were unwilling to back down, they considered or adopted policies to prepare for a conflict they thought might be imminent.

Alternatively, consider the Japanese strategic calculus in 1941. In November of that year, the United States demanded that Japan withdraw from China. This amounted to "surrendering her position as a power in the Far East."[28] Japan was unwilling to accede to U.S. demands, and, precisely because it found U.S. threats credible, decided to take radical action to prepare for the coming conflict. In the hope of demonstrating its resolve to resist, and of engaging in only a limited war with the United States, Japan opted to destroy the offensive capability of the U.S. fleet at Pearl Harbor.[29]

The history of international relations is full of examples of states responding to diplomatic pressure with actions that go well beyond a simple refusal to comply with a demand. In fact, where important questions of security are concerned, simply declining to comply with demands is likely the exception rather than the rule. The Japanese response to U.S., British, and Dutch policy in 1941 was exceptional only in scale and decisiveness.

In response to Austrian threats during the Crimean War, Russia took actions it would not have taken otherwise. These included colluding with France and Sardinia to strip Austria of northern Italy, tipping the balance in Germany in favor of Prussia, permitting the revolution in Hungary that resulted in the Austro-Hungarian Ausgleich (instead of assisting in suppressing the Hungarians as Russia had before the war), and declining to renew generous offers of Russo-Austrian cooperation in the Balkans, leading to drastically increased security competition in the region between the two countries.[30] More speculatively, but with some justice, the historian Norman Rich argues that the Austrian threats during the Crimean War resulted in "a bitter hostility that was to culminate in war in 1914, the destruction of both imperial houses, and the liquidation of the Habsburg Empire."[31]

A particularly common response to dissatisfaction with another country's conduct of foreign policy and its perceived hostile intentions is the realigning of alliance commitments. In 1864, for instance, Napoleon III wished to use a European conference to revise the post–Napoleonic Wars settlement of 1815. When Britain, with which France was closely aligned, refused to support a conference, Napoleon was explicit: "So it seems we shall have no Congress. Well! I shall have to change my alliances." With that, the alignment between the two countries ended.[32]

The danger of a great power realignment resulting from a general breach in Russo-German relations is also the explanation for the information conveyed by Germany's statements in 1876 discussed above. Both sides understood that even a tacit threat from Germany might lead Russia to form an alliance with France, Germany's arch rival since the Franco-Prussian war six years before. The fact that Germany understood the danger, and yet chose to tacitly threaten anyway, meant that Germany had effectively communicated its resolve to defend Austria-Hungary.[33]

To fix ideas, and to see the relationship between preparation for conflict and communication, consider the following stylized examples of international contexts in which costless communication is possible. The model presented in the next section is designed to represent these situations in a stylized way. As these narratives make clear, signaling dynamics can be similar in seemingly diverse strategic contexts.

***Example 1* (External Balancing).** A conflict arises between two states, a Deterrer and a Target, over a specific issue. In order to get its way, the Deterrer would like to signal its willingness to go to war over the issue to the Target. The Deterrer knows, however, that if the issue is particularly important to the Target, the latter may form an alliance with a third country in order to get its way or prepare for a possible conflict. The Target would prefer not to make the concessions to the third state required to get an agreement, and the new alliance is also likely to have a negative effect on the Deterrer's security position, especially if the third country is already hostile to the Deterrer. Threat-making therefore has both an advantage and a drawback. The advantage is the increased likelihood that the Target state will concede the issue to the Deterrer; the drawback is the possibility that the Target will "balance" against the Deterrer by forming a hostile alliance. As has been mentioned, this was the principal concern in German relations with Russia after the Franco-Prussian war. Deterrers for whom the issue is not sufficiently important are unwilling to incur the risk of such a breach in relations by making a threat. When the Target state observes a threat, therefore, it learns the issue is relatively important to the threatening state.

***Example 2* (Internal Balancing).** As the name implies, this scenario is similar to the last, except that the principle strategic option of the Target (the threatening state's principal concern) is to transfer resources to its military sector in order, one day, to resist the demands of the threatening state. The timing of China's decision in the 1950s to devote enormous diplomatic and material resources to the pursuit of nuclear weapons, for instance, may have been partly a result of U.S. threats in the first and second Taiwan Strait Crises. Because

[27] Cited in Press (2005, 5); italics added.
[28] Feis (1950, 327).
[29] See George (1991, 19) and Russett (1967).
[30] For a detailed comparison and analysis of Russian foreign policy before and after the Austrian threats, see Trager (2007, chap. 2).
[31] Rich (1965, 123).
[32] Mosse (1958, 142).
[33] For an analysis of this case, see Trager (2007, chap. 3). For a related argument on the follow-on effects of alliance realignments, see Healy and Stein (1973).

such arms production alters the future bargaining relationship between the nations, internal balancing will often constitute a significant long-term drawback to threat-making.

***Example 3* (First Strike).** If a Deterrer threatens a Target, the military and strategic context may be such that if the Target is unwilling to back down and believes the Deterring state is also sufficiently unlikely to back down, the Target's best option is to strike first. This was the situation for Japan in 1941. It was also a worry for U.S. President Kennedy during the Cuban missile crisis. He recognized that a U.S. threat to destroy the missiles in Cuba in four days could result in a Soviet threat to take action in three days, as well as further escalations that could result in nuclear war in that time frame.[34] In such military-strategic contexts, therefore, threat-making once again involves a trade-off similar to that in Examples 1 and 2. Threats increase the chance that a state gets what it wants with respect to the issue at hand, but can also create a danger that the threatened side will begin an unwanted military conflict. Once again, the willingness of the Deterrer to incur such a risk can cause the threat to convey information to the Target.

***Example 4* (Too Costly Deterrence).** Faced with a threat, the Target might consider adopting policies that would deter the threatening state from attacking. Such activities, for instance mobilizing forces on a border, may be too costly to sustain for long. Rather than maintain a high level of preparation, therefore, the Target state might prefer to go to war. Powell (1993) analyzes a strategic context that leads to a similar dynamic. From the perspective of diplomatic signaling, if we think of the Target's decision as preceding the Deterrer's decision to go to war or back down, the danger of such an outcome plays a similar role to the risk of a first strike in Example 3.

***Example 5* (Resource Drain).** Other reactions that states may have to threats also have long-term consequences for the threatening state. As the Soviet Union did following the Vienna meeting, for instance, a state may choose to drain the resources of another state in order to get its way on a particular issue. As in Examples 1–4, the risk that a threatened state will adopt such a course provides a disincentive to less resolved states to signal their willingness to engage in conflict over such a contentious issue, making communication possible.

***Example 6* (Mobilization).** When a threatened state declines to back down and believes conflict likely, it may elect to mobilize its troops. With the cost of the mobilization paid, the choice to go to war looks more attractive to the mobilized state than it had previously.[35] If we allow for the possibility that the mobilized state will go to war, we once again have a tradeoff for states that consider threat-making, and this again results in the possibility of informative signals.

[34] May and Zelikow (2002, 43–4).
[35] Slanchev (2005).

In each of these narratives, the decisions the Target of a threat takes when it means to resist the deterring (or compelling) state's demands have a negative impact on the Deterrer's utility. If the Target chooses either internal or external balancing, the Deterrer's utility will be negatively affected even if it chooses not to go to war. This is true when the increased capabilities of the Target increase the likelihood that the Target will decide to go to war itself, and may also be true when the Target will not contemplate war in the near term. In the context of Russo–German relations in 1876, Bismarck was explicit on this point. Even a tacit German threat to Russia, he argued in one foreign policy circular, "could induce [the Tsar] to conclude flawed resolutions and alliances that would be very disadvantageous for both sides."[36] A Franco-Russian alliance would have had consequences for Germany whether or not the Germans backed down in 1876.[37]

Internal and external balancing on the part of a Target negatively affects the Deterrer's expected utility from peace because it changes the balance of power between them. If the two countries are involved in another crisis in the future, the weaker relative position of the Deterrer will usually mean that it is less likely to get its way and more likely to fare badly if conflict should actually break out. Thus, if future crises are of the sort described here or, for example, in Fearon (1994, 1997, 1998), Schultz (1998, 2001) or Slanchev (2005), the Deterrer's expected future utility decreases in the capabilities of its adversary.[38] Although there may be some contexts in which states are indifferent to the increasing capabilities of an adversary, they will more often view such developments with understandable concern.

In the model described below, we allow for the possibility that the Target of a threat, having made costly preparations for war, might choose to begin one. This results in a concrete risk to the coercing state of appearing to menace the other state. Even if the coercing state backs down, the other state may make preparations for war and choose to fight one. More generally, however, we might think of the preparations of the Target negatively affecting the coercing state's utility because of the effect of preparations on the balance of power, even when the coercing state backs down and the sides are not in immediate conflict.

[36] *Grosse Politik*, v. II, p. 37.
[37] Note that Russia elected not to attack or explicitly threaten Austria-Hungary and that Russia also did not pursue an alliance with France at this time. The Franco-Russian alliance came about 15 years later, after much intervening history, including Russia's declared frustration with German policy in the Congress of Berlin, the Austro-German alliance of 1879 (directed partly, and only partly secretly, against Russia), and the German failure to renew the Reinsurance Treaty with Russia in 1890.
[38] Kydd (2005) is a partial exception because increased relative capabilities of an adversary may on occasion make it more trusting and thereby result in a net improvement of a threatening state's expected utility. We could also consider multistate interactions in which the increased capabilities of an adversary are beneficial, but the reverse is of course the more usual case.

COMMUNICATION IN ANARCHY

We now turn to a formal model of coercion in which we allow for the possibility that threatened states may decide to take actions that go beyond standing firm or backing down. In the model, there are two players, a Deterrer or threatening state and a Target or threatened state. The players are indexed by $i \in I \equiv \{d, t\}$.

The model below is different from previous models in the international relations literature in two important respects. First, the Target can chose to prepare for conflict when it believes it is sufficiently likely, or reorient its foreign policy in other ways that are hostile to the interests of the threatening state. If the Deterrer takes a threatening posture, and if the Target's costs of war are sufficiently low relative to the importance of the issue in question, the Target may be convinced to take these sorts of measures. Second, in addition to the Deterrer having the option to attack the Target, the Target has the opportunity to attack the Deterrer. Having prepared for war, the Target may opt for it even if the Deterrer does not. Because the Target's actions may lead to an outcome that is even worse for an unresolved Deterrer than the peaceful outcome where the Target is undeterred, a resolved Deterrer has an incentive to make threats vis-a-vis a particular issue that less resolved states would be unwilling to make. As I show below, this dynamic causes threats to be meaningful even though no direct cost is associated with making them. One general implication of the model is therefore that if states respond at all to perceptions of other states' intentions in formulating foreign policy, then a verbal threat made by one state against another, whether in public or private, can convey information.

Following other models in the international relations literature, we shall suppose there is a bargaining space $X \equiv [0, 1]$ such that the Deterrer prefers outcomes closer to 1 and the Target outcomes closer to 0. The status quo at the beginning of the game is $s \in X$. Players have von Neumann–Morgenstern utility functions defined over outcomes in the bargaining space, $x \in X$, and whatever costs of fighting and preparation for conflict the players pay, c_i. Specifically;

$$u_d(x) = x - c_d$$

and

$$u_t(x) = 1 - x - c_t.$$

Thus, players have risk-neutral preferences over outcomes in the bargaining space. We assume this because it simplifies the exposition without substantively altering the key points of the analysis. The costs players pay in the game, c_d and c_t, will be defined below as functions of other variables to reflect the outcomes of player actions during the game, so that, for instance, the players do not pay a cost of conflict when no war is fought. (They may still pay a cost of preparing for conflict, however.)

Figure 1 depicts the stages of the game. It begins with the Deterrer's attempt to influence the Target by sending a costless signal. Whatever message the Deterrer sends, the game that follows is precisely the same; action sets, the sequence of moves, and player utilities are all unchanged. If the initial communication stage is important at all, therefore, it is only because the Deterrer's message affects the Target's *beliefs* about the Deterrer's intentions. This is possible in equilibrium when the Deterrer is aware of the conclusions the Target will draw from a particular messaging strategy of the Deterrer (and thus also aware of how these conclusions will likely affect its actions), and yet the Deterrer's best option is nevertheless to use that very messaging strategy. We restrict attention to the Deterrer's decision to threaten or not, which we represent by the message $m \in M \equiv \{0, 1\}$. We shall interpret $m = 1$ as a promise to take violent action if the Target takes a particular action, as well as a promise not to take violent action in the event the Target cooperates.

FIGURE 1. Stages of the Game

	Deterrer
First stage	Makes a costless threat or acquiesces
	Target
Second stage	Decides (1) whether to take the action in question and (2) whether to prepare for conflict
	Deterrer and Target
Third stage	Each decides whether or not to go to war

In the second stage, the Target has two decisions to make: whether to take the action in question or not (choosing $a_1 \in A_1 \equiv \{0, 1\}$), and whether to prepare for conflict (choosing $a_2 \in A_2 \equiv \{0, 1\}$). If the Target takes the action, setting $a_1 = 1$, it unilaterally moves the status quo to $s - \epsilon$, toward its ideal point, where $\epsilon \in (0, s]$. If the sides remain at peace, this will then be the bargaining outcome. Then, in the third stage, the Deterrer and Target decide whether or not to go to war by choosing $r_i \in R_i \equiv \{0, 1\}$, where 1 represents conflict initiation. First, the Deterrer chooses r_d. If $r_d = 1$, conflict occurs; if not, the Target chooses r_t. War occurs if either side opts to begin one; both sides must choose peace to obtain that outcome. We let r represent an indicator variable that equals 1 when the sides go to war and 0 otherwise. Thus, $r = 1$ if and only if $r_i = 1$ for some i.

If the sides should fight a war, the Deterrer will win the war with common knowledge probability $p(a_2)$, and is then able to choose its ideal outcome in the bargaining range X. Similarly, if the Target should win the war, it may choose its ideal bargaining outcome. We assume that one of the two sides will win the war. When the Target prepares, the chances that it wins a war may increase: $1 - p(1) \geq 1 - p(0) \Leftrightarrow p(0) \geq p(1)$.

We shall model the Target's costs of war and preparation as consisting of several components. If the sides go to war, the Target's war costs are $\eta_t \in [\underline{\eta}_t, \overline{\eta}_t] \equiv \Xi_t$, where $\underline{\eta}_t > 0$ and η_t is private information of the Target. If the Target chooses to prepare, it incurs some preparation costs, $k_t \geq 0$, whether or not a war is fought,

but reduces its costs of conflict by $\beta_t \in [0, \underline{\eta}_t)$. Thus, preparations imply an increase in the sunk costs that the Target pays whether or not it goes to war and a decrease in the variable costs associated with the conflict itself. (We make no assumption about the net effect of preparations on the overall cost of conflict.) Thus,

$$c_t(a_2, r, \eta_t) = k_t a_2 + r(\eta_t - \beta_t a_2).$$

We take a similar approach to modeling the Deterrer's costs, but we suppose the Deterrer has already made any relevant preparations for conflict and do not model the Deterrer's choice of preparations explicitly. For simplicity, we suppose the Target's preparations do not affect the Deterrer's costs of conflict—only its probability of victory. Thus,

$$c_d(r, \eta_d) = r\eta_d,$$

where $\eta_d \in [\underline{\eta}_d, \bar{\eta}_d] \equiv \Xi_d$, $\underline{\eta}_d > 0$, and η_d is private information of the Deterrer. Both sources of private information are independently distributed according to the continuous, strictly increasing, common knowledge distribution functions Φ_{η_i}.

Thus, the Deterrer's utility depends on the bargaining outcome, whether the players go to war, and the Deterrer's type, whereas the Target's utility depends on those same factors and also on whether or not the Target chooses to prepare. Therefore, we shall write the players' utility functions as $u_d(x, r, \eta_d) : X \times R \times \Xi_d \to \mathbb{R}$ and $u_t(x, r, \eta_t, a_2) : X \times R \times \Xi_t \times A_2 \to \mathbb{R}$.[39] Substituting the cost functions into the player utility functions yields the following utilities for peace:

$$u_d(s, 0, \eta_d) = s - \epsilon a_1$$

$$u_t(s, 0, \eta_t, a_2) = 1 - s + \epsilon a_1 - k_t a_2.$$

Similarly, the players' expected utilities for war are[40]

$$Eu_d(r = 1 \mid a_2, \eta_d) = p(a_2) - c_d(1, \eta_d) = p(a_2) - \eta_d$$

$$\begin{aligned} Eu_t(r = 1 \mid a_2, \eta_t) &= 1 - p(a_2) - c_t(a_2, 1, \eta_t) \\ &= 1 - p(a_2) - \eta_t + \beta_t a_2 - k_t a_2. \end{aligned}$$

The formal structure of the game is shown in Figure 2. Whichever choice the Deterrer makes in the first stage, the Target has the four Stage 2 options shown in the figure. Whichever of these the Target opts for, the same Stage 3 structure shown follows. As the figure shows, the players' utilities over the Stage 3 terminal node outcomes depend on the Target's Stage 2 choice. The full game tree, therefore, has 24 terminal nodes (2 Deterrer options in the first stage × 4 Target options in the second stage × 3 terminal nodes in each Stage 3 branch).

To make the signaling problem interesting, we shall make several assumptions about payoffs. First, we assume the Deterrer prefers not to go to war when the Target complies with its wishes by not taking the action, and that there is a possibility the Deterrer would be willing to fight if the Target did not prepare and took the action in question and a possibility the Deterrer would not be willing to fight in this case. This is equivalent to

$$s > p(0) - c_d(1, \underline{\eta}_d) > s - \epsilon > p(0) - c_d(1, \bar{\eta}_d). \quad \textbf{(1)}$$

Second, we assume the Target prefers peace when it takes the action in question and does not prepare to its most preferred war outcome and that there is a possibility the Target would prefer a prepared war to accepting the initial status quo and a possibility the Target prefers the initial status quo to a prepared war. This is equivalent to

$$1 - s + \epsilon > 1 - p(1) - c_t(1, 1, \underline{\eta}_t) > 1 - s > 1 - p(1) - c_t(1, 1, \bar{\eta}_t). \quad \textbf{(2)}$$

Similarly to other models of coercion, these assumptions imply that the Deterrer's most preferred outcome occurs when the Target complies with the Deterrer's wishes by not taking the action in question (and also not attacking the Deterrer). The Target's most preferred outcome occurs when it takes the action in question, does not prepare for conflict, and the Deterrer does not attack. Players are willing to go to war when they consider the costs of war to be low relative to their evaluation of the issues at stake and their chances of victory.[41] An example of parameters satisfying these assumptions is shown in Figure 3. The figure also represents the model in terms similar to those used in Fearon (1995) and can usefully be compared to Figure 1 of that paper.

The preparations of the Target may or may not have an effect on the Target's costs of conflict and the players' probability of victory. We shall say that preparations are *effective* if and only if $k_t > 0$, $p(0) > p(1)$, and $[(1 - p(1)) - (1 - p(0))] > k_t - \beta_t$. The last condition states that the Target prefers a prepared to an unprepared war, or in other words, that the overall increase in the Target's costs as a result of preparation (*if* preparing does involve a net increase in costs) must be outweighed by the benefit of the increased likelihood of victory. We shall say preparations are *ineffective* if $k_t = \beta_t = 0$ and $p(0) = p(1)$.[42] Note that in Figure 3, preparations are effective.

[39] Note that although the Target's choice of preparation does not affect the Deterrer's utility directly, it does affect the likelihood of outcomes (victory and defeat) over which the Deterrer has different preferences.

[40] So long as player utility functions are bounded, we can derive these expected utilities without assuming risk neutrality over bargaining outcomes by setting $u_d(1, 0, \eta_d) = u_t(0, 0, \eta_t, 0) = 1$ and $u_d(0, 0, \eta_d) = u_t(1, 0, \eta_t, 0) = 0$ without loss of generality. Because the sides' war utilities do not depend on which side chooses war in the third stage, the order of player choices in this stage will have no substantive impact on the analysis.

[41] Here, players are uncertain about each other's costs of conflict. We might also consider a model in which players are uncertain about the importance of the issue in question to the other side. A model of this form, which generates results similar to those presented below, can be found in Trager (2007, chap. 3).

[42] The effective/ineffective distinction does not exhaust all regions of the parameter space. We shall not analyze cases where preparations impact the probability of victory but are not costly or are costly but do not affect the chances of victory.

FIGURE 2. Formal Structure

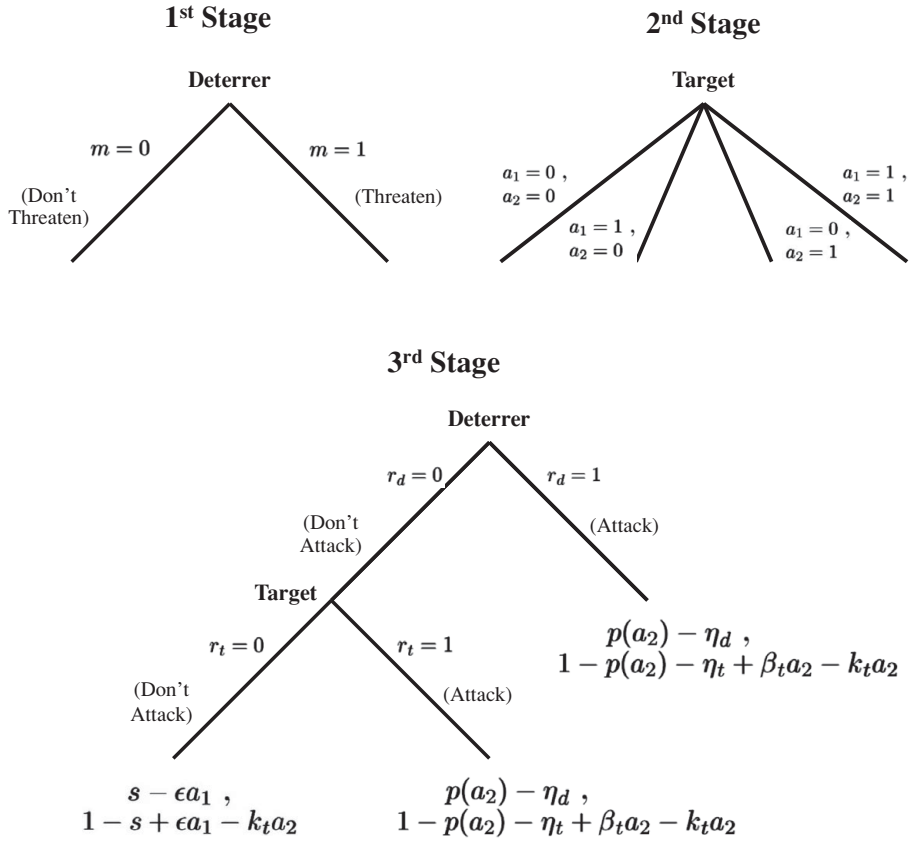
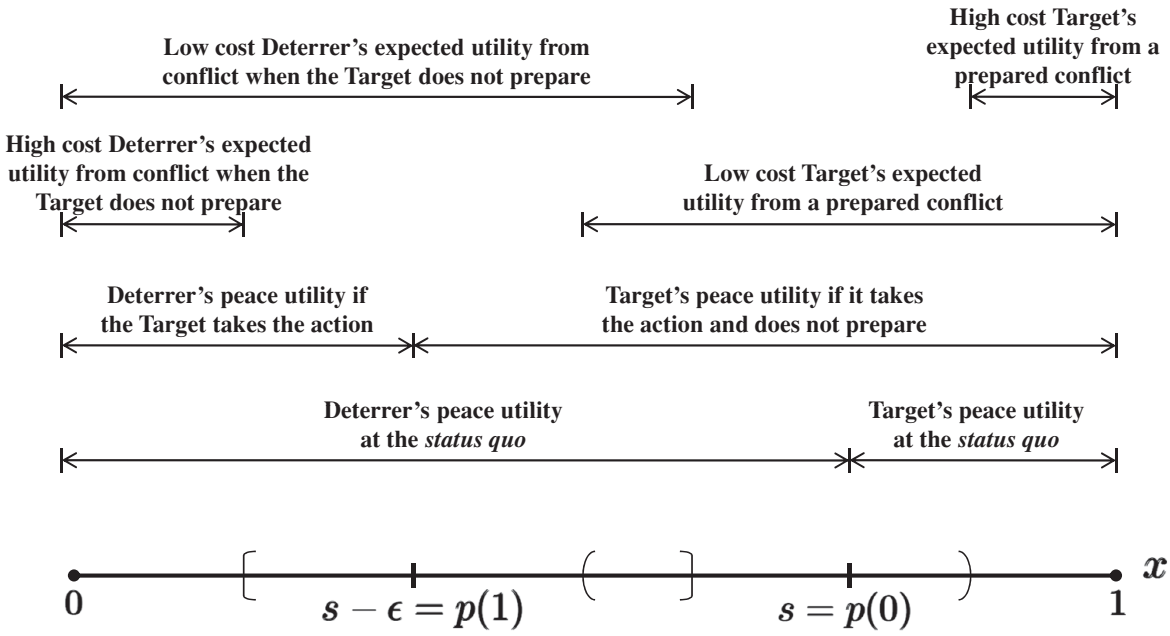


FIGURE 3. Player Utilities Satisfying Model Assumptions



Equilibria with Ineffective Preparations

We can now ask whether, in such a strategic context, the Target can learn from the costless signals sent by the Deterrer. Commonly, in cheap talk games, signalers can convey information that does not influence the course of events. If communicated information will not influence actions, it makes little difference to the receiver of the information whether or not he or she believes it to be true. In these equilibria, signals are *informative* but not *influential*. We shall focus on whether equilibria with influential signals can exist. In such equilibria, there are signals that increase the probability that the Target both does not take the action in question ($a_1 = 0$) and does not attack the Deterrer ($r_t = 0$). Let $q_t(m) = \Pr(a_1 = 0, r_t = 0 \mid m)$ be the probability the Target's strategy satisfies $a_1 = 0, r_t = 0$ given signal m induced by the players' strategies and beliefs in a particular equilibrium.

Definition. An equilibrium is influential if there exist signals $m' \neq m''$ played with positive probability in the equilibrium such that $q_t(m') > q_t(m'')$.

If the Target's preparations do not affect its probability of victory or the players' costs of conflict, the Deterrer's costless signals cannot convey information to the Target in a way that influences the Target's actions; no influential equilibrium exists. If preparations have no effect on player utilities over outcomes, then this model is very similar to others in the international relations literature. In such cases, the reason informative signaling is impossible is well understood: there is a benefit to being thought willing to fight and no drawback.[43] Therefore, the Deterrer always tries to send the signal that will most convince the Target of its resolve, but because that signal is sent in every case regardless of whether it corresponds to the truth of the matter, the Target learns nothing from it. Proposition 1 expresses this formally.

Proposition 1. *If preparations are ineffective, no influential perfect Bayesian equilibrium exists (in pure strategies).*

Influential Equilibria

When preparations have an effect, in some cases influential equilibria exist in the model. To understand how these equilibria function intuitively, suppose the Deterrer makes a threat and the Target learns from this that the Deterrer is more likely willing to go to war over the issue in question than the Target had previously thought. (We shall see that, in this equilibrium, this supposition is correct.) Realizing that the Deterrer is more likely to follow through, the Target is more likely to decide that the issue is not worth risking a fight over and to back down. If the Target's costs of war are sufficiently low relative to the value it places on the issue, however, the Target may decide not to back down and to make costly preparations for war. The Target's decision may be a difficult one because on the one hand, preparations such as alliances, building arms, or striking first may provide additional security, but on the other hand, such actions carry their own additional costs. Because the Target's preparations affect its own calculations about the relative benefits of war and peace, having prepared, it may decide to attack the Deterrer.

In the first stage, in deciding whether or not to threaten, the Deterrer understands these dynamics. If the Deterrer's costs of war are high relative to the value it places on the issue in question, then even though by threatening it stands a better chance of getting its way, it will not be willing to make a threat. This benefit to threatening does not outweigh the increased risk of a breach in relations with the Target and the attendant increased risk of a costly conflict. Thus, the Deterrer will only be willing to threaten if it is relatively highly resolved. This, in turn, implies that our supposition that the Target learns from the threat is correct.[44]

A principal obstacle to influential cheap talk signals is the incentive that unresolved types have to mimic the signals sent by resolved types. In other models in the literature, unresolved types have every incentive to mimic because they can always back down and ensure themselves of an outcome they like just as well as the one in which they make no threat. In the model described here, however, there may be a danger that the Target will respond to a threat by preparing for war and, having done so, launch a strike of its own. When this is possible, low-resolve types face a risk in misrepresenting their levels of resolve by mimicking the behavior of high-resolve types.[45]

[43] The logic here is only slightly more complicated because, unlike the situation in most models in the literature, the Target also decides whether or not to go to war later on. The proof of Proposition 1 must also demonstrate, therefore, that no influential equilibrium exists in which resolved, threatened Targets decline to take the action in question, thereby committing themselves to fighting a war. If such a dynamic were possible, less resolved Deterrers might decline to threaten, resulting in influential signaling. As the proof of Proposition 1 in the Appendix demonstrates, however, this cannot occur.

[44] As described, the logic may appear circular. But note that it is circular only in the way that the logic of any truly strategic Nash equilibrium (where players' optimal actions are dependent upon the actions of other players) must be circular.

[45] In the model, the Target can take an action that shifts the bargaining outcome in its favor by ϵ or not take that action. As a result, the Target sometimes prefers to go to war because the shift in the bargaining outcome is not sufficient to cause a fully prepared Target to prefer peace. If it were feasible for the Target to take an action unilaterally that moved the bargaining outcome sufficiently further toward the Target's ideal point following preparations for conflict, however, and if we allowed for this possibility in the model, then the Target would never choose war in equilibrium and influential signaling would be impossible. The assumption of a fixed action negotiated over by the states is probably reasonable in some cases and not in others. Even if we were to change the model to allow for an increase in ϵ following preparations, however, influential signaling would still be possible if we relaxed the assumption of risk neutrality over bargaining outcomes. More generally, for alternate bargaining protocols, influential signaling would be possible as long as Target types with higher expected values for war take actions that imply a strictly higher likelihood of conflict. This is an assumption or a result

For this logic to operate, several conditions must be met. On the one hand, Deterrers cannot have so much to gain from threatening that even the least resolved Deterrers would be willing to do it. If the Target is thought too likely to make concessions, for instance, the Deterrer will always find it preferable to try its luck with a threat. Similarly, and as we saw above, if the preparations of the Target are not effective enough, there will be no disincentive to making threats all the time.

On the other hand, resolved Deterrers must have enough to gain from threatening. If the Target is thought too likely to make preparations, if those preparations have too large an effect on the balance of power, or if there is not a high enough likelihood that the Target will concede the issue, then the Deterrer will prefer to mislead the Target about its true level of resolve. States willing to go to war will see it in their interest to convince Targets that they are not, in order to catch the latter unprepared. This too can make informative signaling impossible.

Proposition 2 establishes one set of sufficient conditions for influential signaling. So long as preparations are minimally effective and both sides are not too certain about the resolve or irresolve of the other side, an influential perfect Bayesian equilibrium exists when the sunk costs of preparation, k_t, are at least as great as the value of the issue in contention, ϵ. The significance of this last condition is that it ensures that when the Target finds war an unattractive option, it prefers to comply with the Deterrer's demand rather than to make preparations in order to deter an attack from the Deterrer while declining to comply. If Targets preferred noncompliance with preparations to compliance and if the impact of Target preparations on the probability of victory were sufficient to prevent Deterrers from ever attacking, then Deterrers would have no incentive to signal resolve. They would have nothing to gain from doing so, and an influential equilibrium would not exist.

Proposition 2. *When preparations are effective, for some set of beliefs Φ_{η_t}, an influential perfect Bayesian equilibrium exists if $k_t \geq \epsilon$.*

As the discussion above makes clear, the sufficient conditions provided in Proposition 2 are not necessary for the existence of influential equilibria. In particular, if there is a sufficiently high likelihood that the Deterrer would be willing to fight even against a prepared Target, influential equilibria exist in some cases when $k_t < \epsilon$. The reason is that the willingness of the Deterrer to fight ensures that the Target might sometimes be unwilling to risk war and thus would prefer to comply with the demand rather than preparing and declining to comply. Thus, by demonstrating its resolve, the Deterrer would have something to gain, enabling influential signaling.

To better understand how signaling operates in the model, we now consider a particular parameterization. Let us suppose that if the sides fight a war, there is a $p(0) = 50\%$ chance that each side wins when the Target has not made preparations. The status quo in the issue space, s, favors the Deterrer at 0.6, but if the Target takes the action in question, it unilaterally shifts the status quo by $\epsilon = 0.2$ to 0.4. If the Target signs an alliance with a third country, mobilizes its troops, and initiates a dramatic armaments program, the probability that the Deterrer is victorious in war decreases to $p(1) = 30\%$. The Deterrer's costs of war range from small to quite large: η_d is uniformly distributed over $[0.15, 0.75]$. We shall suppose that the Target's costs of war are very high when it does not prepare, but similar although somewhat less than the Deterrer's costs when the Target does prepare: $\beta_t = 0.585$, η_t is distributed uniformly over $[0.6, 0.85]$, and the Target's fixed costs of making preparations are $k_t = 0.2$. Thus, if the Target prepares, the additional costs it pays as a result of the conflict $(\eta_t - \beta_t)$ range from 0.015 to 0.265.

In such a world, there is an equilibrium in which the Deterrer sends influential, costless signals of its resolve. Before the Target observes the Deterrer's choice to threaten or not, it believes there is only a 14% chance that the Deterrer would be willing to fight over the issue. But the Target also knows that if it observes a threat, the Deterrer is willing to take a risk of war. This is because the Target knows that the Deterrer believes there is a 34% chance that the Target's war costs once it prepares $(\eta_t - \beta_t)$ are in the range $[0.015, 0.1)$. If the Target's costs of war are in this lower range, it prefers a prepared conflict to accepting the status quo. Its optimal response to a threat from the Deterrer is to make maximal preparations for war. There are fixed costs of $k_t = 0.2$ associated with these preparations that the Target pays whether or not the countries go to war. Having paid this cost, the Target's expected utility from conflict is $1 - p(1) - k_t + \beta_t - \eta_t = 1 - 0.3 - 0.2 + 0.585 - \eta_t = 1.085 - \eta_t$. For low-cost Targets (e.g., $\eta_t = 0.6$), this is greater than its utility when conflict is avoided (even when it gets its way with respect to the issue in question), which is $1 - s + \epsilon - k_t = 0.4$. Thus, if the Target is threatened and if it is highly resolved (has relatively low costs of war), the Target will make preparations and go to war with the Deterrer.

Optimal behavior by the Target therefore implies that threat-making by the Deterrer entails a risk of conflict. For this reason, the Deterrer would only be willing to make a threat when its privately known costs of conflict (η_d) were less than 0.42. Thus, the Deterrer threatens only 46% of the time. Because the Deterrer is willing to go to war against an undeterred and unprepared Target only when $\eta_d < 0.235$, the Target knows for sure when the Deterrer declines to threaten that it will also decline to initiate conflict if the Target takes the action in question and does not prepare. On the other hand, if the Deterrer does make a threat, the Target will believe there is a 31% chance that the Deterrer would initiate conflict if the Target did not back down and did not prepare. This, in turn, causes the

of nearly all crisis bargaining models. See Banks (1990) in particular, which demonstrates in a fairly general setting that the probability of war is at least weakly increasing in the expected conflict utility of an informed player.

equilibrium to be influential: the probability that the Target backs down is 0% when the Deterrer declines to threaten, and 66% when the Deterrer does.

The reason the equilibrium is influential is that the Target learns from the Deterrer's costless statements. The Deterrer's initial belief is that there is a 14% chance the Deterrer would be willing to go to war over the issue. When the Deterrer threatens, that probability more than doubles to 31%. When the Deterrer declines to threaten, that probability falls to 0%.

The difference from other deterrence models where the Deterrer's statements could not convey information is the possibility that the Target will take an action that negatively affects the Deterrer's utility whether or not the Deterrer later opts for war itself. In deciding whether or not to threaten, therefore, the Deterrer considers the tradeoff described in the stylized examples of the previous section. If it declines to threaten, it reveals itself as a low type, ensuring that the Target will not concede the issue in question. Less resolved Deterrers, those that have a high relative cost of conflict, are nevertheless willing to make this choice because of the risk of unwanted conflict that threat-making entails. On the other hand, the possibility the Target will back down without the need for conflict makes it worth it for resolved Deterrers to apprise Targets of their intentions by making a threat.

Noninfluential Equilibria

Even when preparations are effective, influential equilibria may not exist. When Deterrers risk too little by making threats, they will not be able to influence Targets. When they risk too much, they will not be willing to make threats at all. In either case, influential signaling is impossible.

Consider, for instance, Deterrers that are known to have relatively low costs of conflict, when there is a low upper bound on Deterrer cost types ($\bar{\eta}_d$). In such cases, Deterrers risk relatively little by threatening. So long as there is a high enough likelihood that the Target has relatively high costs of war and therefore will be willing to back down, the Deterrer will always prefer to make a threat. Thus, in such cases, signaling will be impossible for the traditional reason: low-resolve types (high–cost of war types) prefer to mimic high-resolve (low-cost) types.

Conversely, if the Target is thought very likely to be a relatively low-cost type and thus likely to prefer war to complying with the Deterrer's demand, signaling will again be impossible, but for a different reason. Instead of an incentive to misrepresent itself as resolved even when it is not, the Deterrer would have an incentive to misrepresent itself as unresolved even when it is. Rather than threaten in the hope the Target would improbably be willing to comply, resolved Deterrers would prefer to pretend to be unresolved in hopes of catching the Target unprepared. If the incentive of resolved states to misrepresent themselves as unresolved is strong enough, influential signaling will again be impossible.

When resolved states have such incentives to mimic unresolved states, wars will occur that both sides would have preferred to avoid—and that could have been avoided if the Deterrer had sent a different signal. We shall refer to these as "diplomatically avoidable wars." The intuition for why such wars occur is simple. Consider a Deterrer willing to fight an unprepared Target unless the Target makes a concession. If the Deterrer makes a threat, there is a chance that the Target will give in, but there is also a chance that the Target will make preparations, which means either that the Deterrer will itself be deterred from attacking or that the Deterrer will have to fight a prepared Target. Therefore, if the probability of Target resolve is too high, the Deterrer will decide not to warn the Target about its intentions. To illustrate these dynamics, we might think of the Israeli attack on the Osirak reactor in Iraq in 1981. If Israel had made a direct threat beforehand to bomb the reactor unless Iraq halted the reactor's construction, Iraq would most likely have mobilized its air defenses rather than comply. Israel would then have had to acquiesce to the Iraqi nuclear program or fight a prepared opponent, in which case Israel would have been better off hiding its intentions in order to catch the Iraqis unprepared.

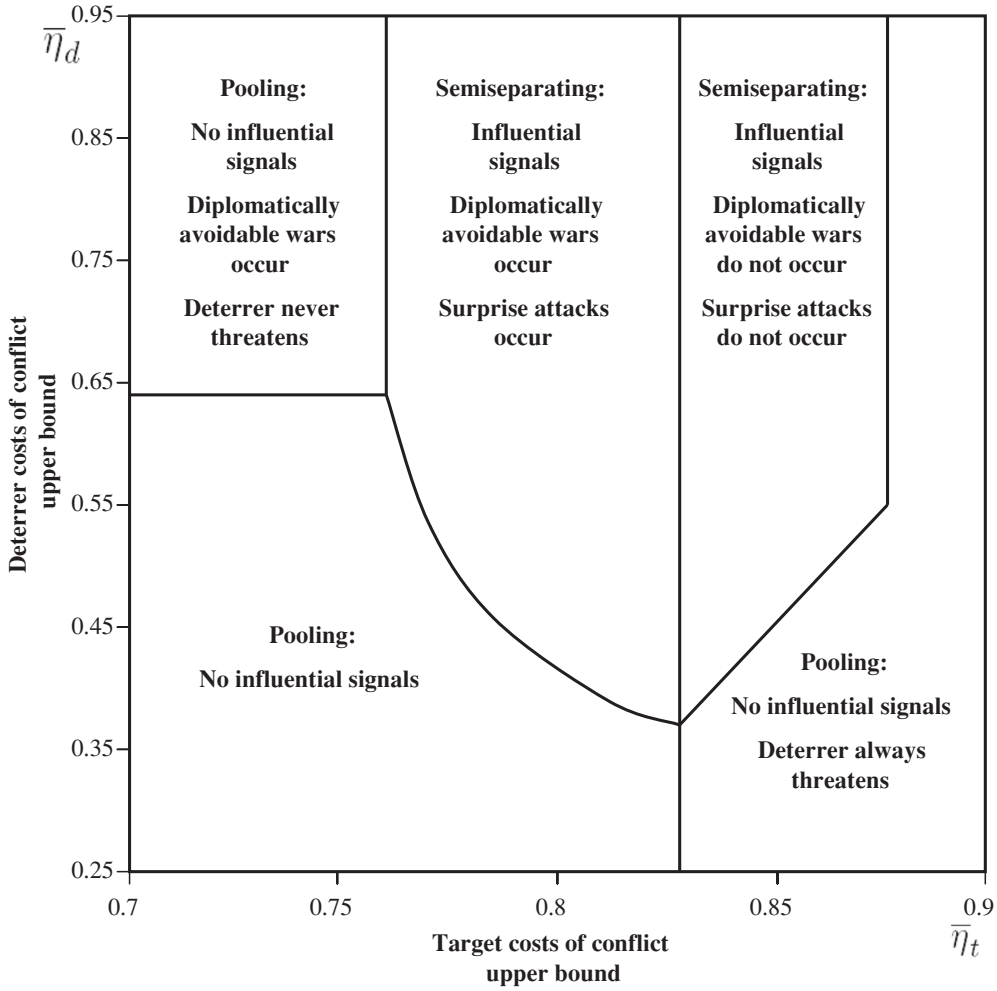
Proposition 3 shows that in any material context, there always exist sets of beliefs of the two sides that imply that diplomatically avoidable wars occur. As the proof in the Appendix makes clear, such wars occur when the Deterrer believes the Target very likely to be highly resolved, whereas the Target believes the Deterrer relatively unlikely to be willing to fight. In such cases, the incentive of resolved Deterrers to mimic unresolved Deterrers completely eliminates the possibility of influential signaling. As we shall see below, however, surprise attacks and their correlate, diplomatically avoidable wars, also occur in some influential equilibria.

Definition. A diplomatically avoidable war is an equilibrium outcome in which (1) the players go to war ($r = 1$), (2) the Deterrer sends some signal m', and (3) the players' strategies imply that had the Deterrer sent a signal $m'' \neq m'$, no war would have occurred ($r = 0$).

Proposition 3. *When preparations are effective, for some set of beliefs Φ_{η}, a noninfluential perfect Bayesian equilibrium exists in which diplomatically avoidable wars occur.*

Another noninfluential equilibrium also exists. Because this is a cheap talk model, an uninformative, "babbling" equilibrium exists in all regions of the parameter space. In this equilibrium, the Target does not interpret the Deterrer's signals as conveying information. Because all Deterrer types are therefore indifferent between signals, the condition that each type is optimizing does not prevent it from sending signals that are uncorrelated with its costs of conflict. In such a case, the Target cannot use Bayes's rule to learn from the Deterrer's signal.

FIGURE 4. Equilibrium Signaling Properties



When influential equilibria exist, empirical analysis of cases stands the best chance of judging whether the influential or babbling equilibrium most closely tracks international reality.[46] There are reasons to suppose the influential equilibrium is a more sensible social equilibrium, however. When the parameters of the model are such that an equilibrium exists in which all Deterrer types willing to go to war would be willing to threaten, for instance, then all Deterrer types that would be willing to go to war (when the Target did not prepare) would prefer to be in the informative rather than the uninformative equilibrium. If we therefore assume such types will attempt to communicate their resolve, then the uninformative equilibrium requires that the very least resolved types (those unwilling to threaten in the informative equilibrium) choose to imitate the behavior of the resolved types by threatening themselves. This, in turn, requires that we presume that these least resolved types attempt to send a signal that they would prefer not to have sent if the signal were believed. If this is unlikely, the informative equilibrium may be more reasonable than the uninformative one.[47]

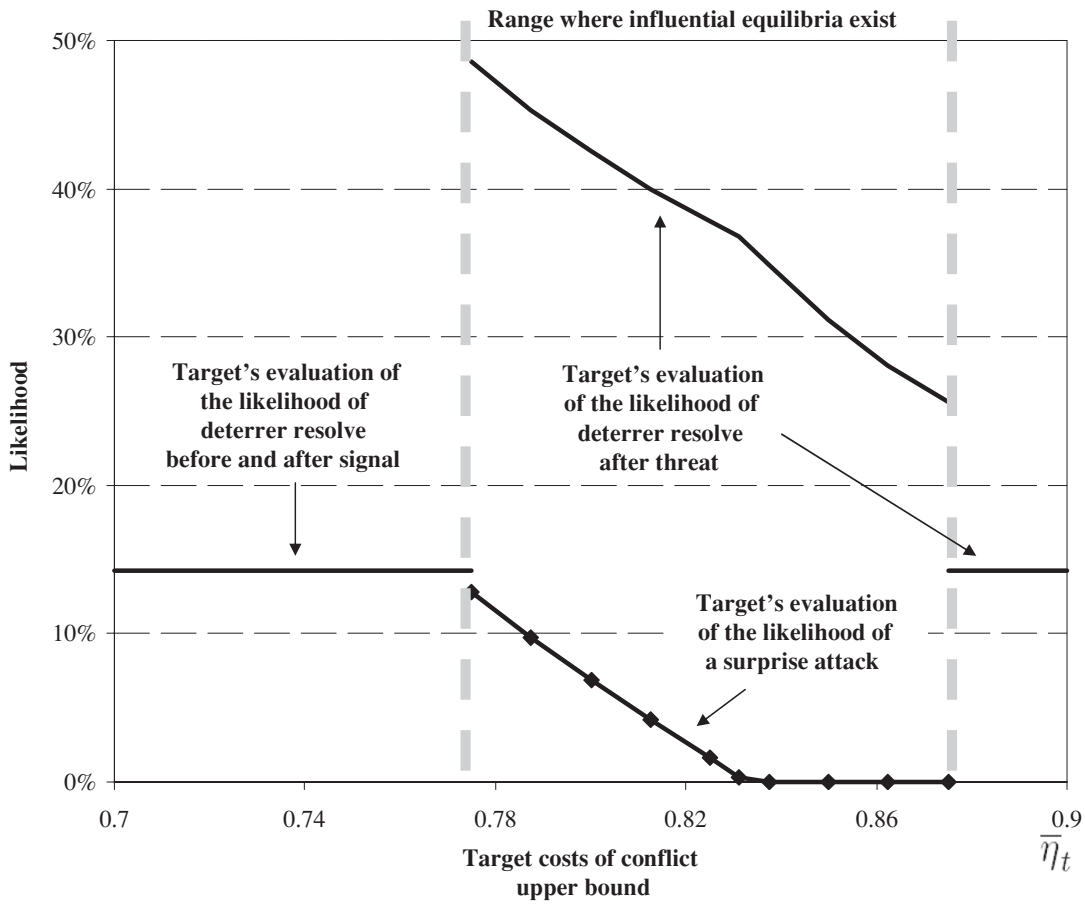
Discussion of Equilibrium Dynamics

We can understand comparative statics in the model by considering the example of an influential equilibrium discussed above. The effect on signaling dynamics of the players' beliefs about each other's willingness to fight is shown in Figure 4. This figure and others that follow were generated through numerical simulations of the model. Recall that in our example, the players' costs of conflict are drawn from uniform distributions over $[\underline{\eta}_i, \overline{\eta}_i]$. We can model shifts in the players' beliefs about each other's resolve by shifts in $\overline{\eta}_d$ and $\overline{\eta}_t$. In Figure 4, $\overline{\eta}_d$ is on the vertical axis and $\overline{\eta}_t$ is on the horizontal axis. Thus, as we move rightward in the Figure,

[46] It may be that process tracing of individual cases stands the best hope of evaluating whether influential or noninfluential equilibria are better descriptions of state behavior. Documentary evidence will sometimes show very clearly, for instance, whether decision maker's beliefs changed after a threat was made.

[47] For other arguments for the reasonableness of informative over uninformative equilibria, see Crawford and Sobel (1982, 1443) and Chen, Kartik, and Sobel (2008).

FIGURE 5. Effectiveness of the Signal and Likelihood of Surprise



the Deterrer is less convinced of the Target's resolve, and as we move upward, the Target is less convinced of the Deterrer's resolve (before a signal is sent).

When the Deterrer is too much or too little convinced of the Target's resolve (very low or very high $\overline{\eta}_t$), or when the Target is too convinced of the Deterrer's resolve (low $\overline{\eta}_d$), influential signaling is impossible. In the middle of the figure are the two ranges where influential signaling is possible, but there are interesting differences in the dynamics of signaling and conflict between these two ranges. In both, the threats of the Deterrer convince the Target that the Deterrer is more likely to be willing to fight over the issue in question than the Target had previously thought. In the rightmost semiseparating equilibrium, however, the Target knows for sure when the Deterrer declines to threaten that the Deterrer will not fight over the issue. In the leftmost semiseparating equilibrium, this is not the case. There, the Deterrer sometimes, but not always, misrepresents itself as *less resolved* than it actually is in order to catch the Target unprepared.

Figure 5 provides us with a closer look at the signaling properties when $\overline{\eta}_d$ is held fixed at 0.75, as in the example of the previous section. $\overline{\eta}_t$ is on the horizontal axis, as in Figure 4. Within the middle range where influential signaling is possible, the more convinced the Deterrer is of the Target's resolve (the lower $\overline{\eta}_t$), the more informative is the Deterrer's signal. When $\overline{\eta}_t$ increases, the probability that the Target complies with the Deterrer's demand increases, which causes the probability that the Deterrer is willing to threaten to increase as well, and the signaling value of a threat to decline. Thus, when $\overline{\eta}_t = 0.775$ a threat causes the Target's evaluation of the likelihood the Deterrer would fight to jump from 14% to 49%. When $\overline{\eta}_t = 0.875$, in contrast, a threat causes the Target's evaluation of Deterrer resolve to change only from 14% to 26%.

When $\overline{\eta}_t$ increases enough, the change in the Target's beliefs as a result of the Deterrer's signal suddenly falls discontinuously to zero as signaling becomes impossible. This can be seen on the right-hand side of Figure 5. The reason is complex and relates to the interaction of a number of factors. As $\overline{\eta}_t$ increases and the Target comes to believe that the Deterrer is less likely to follow through on a threat (because the Target knows that the Deterrer knows that the Target is itself less likely to be willing to fight over the issue and thus the Deterrer has less to lose from making a threat), Target types in the middle of the range of cost types have a new optimal strategy. Even if the Target did find the Deterrer's signal to be somewhat informative, rather than low Target cost types preparing and then fighting and

high Target cost types backing down, Target cost types in the middle of the η_t range would prefer to take the action in question without preparing at all. They find the chance that the Deterrer is willing to fight to be too low to justify the expense of preparation, but neither are they deterred from taking the action in question. As a result of the change in the Target's strategy, the Deterrer would have both less to fear from threatening and less to gain. The former effect dominates the latter, causing the Deterrer to be even more willing to make threats, which in turn would cause even more Target types to prefer taking the action in question without preparation, and so on, with the result that influential signaling is impossible.

When the Deterrer believes it relatively likely that the Target will not back down over the issue, so that $\overline{\eta}_t$ is low, but still in the range where influential signaling possible, surprise attacks (i.e., attacks not preceded by a threat) occur in equilibrium. When the Deterrer's costs of conflict are very low, it prefers not to threaten the Target at all in order to be able to attack the Target when the Target is unprepared. In such cases, as we saw in our analysis of noninfluential equilibria and diplomatically avoidable wars, conflict can occur in equilibrium because the Deterrer is unwilling to risk communicating its resolve. A signal exists that the Deterrer could send that would result in compliance and avoid the need for either side to go to war, but because the Deterrer does not know that the Target would comply with a demand if the Target knew the threat were credible, the Deterrer declines to send the signal.

In traditional crisis bargaining models, war can often occur because the Deterrer does not have the means available to communicate its resolve.[48] As we saw above, in diplomatically avoidable wars, the Deterrer has the ability to communicate but chooses instead to catch the Target unprepared.[49] Instead of the weak Deterrer types mimicking the strong, in such cases, the strong pretend to be weak because they believe the Target is unlikely to back down even if it knew the Deterrer were in earnest.

Unlike the situation in the noninfluential equilibrium of Proposition 3, however, in this example, diplomatically avoidable wars and influential signaling coexist in a single equilibrium. Deterrer's adopt four different approaches, depending on their level of resolve. The very most resolved Deterrers are unwilling to risk Target preparations and choose to use a surprise attack. They will attack an unprepared Target without first attempting to coerce through a threat. The next most resolved Deterrers make threats that they will follow through on if the Target does not acquiesce. Deterrer types that are slightly less resolved than these will make threats they would later be unwilling to prosecute and, finally, the least resolved Deterrer types make no threat at all and later choose not to attack.

Because the likelihood of a surprise attack increases as $\overline{\eta}_t$ decreases, the lower $\overline{\eta}_t$, the less the Target can learn from the Deterrer's decision to *decline* to threaten. When $\overline{\eta}_t = 0.775$, for instance, the Target's evaluation of Deterrer resolve declines only slightly when the Deterrer declines to threaten: from 14% to 13%. When $\overline{\eta}_t = 0.875$, the Target's evaluation of the likelihood of Deterrer resolve falls from 14% to 0%. Thus, when there is a high probability that the Target is highly resolved, the Target learns the most from a threat; but when there is a high probability that the Target is unresolved, it learns the most when the Deterrer declines to threaten.

As can be seen on the left-hand side of Figure 5, at the extreme, when the Deterrer believes the Target very unlikely to back down in response to a threat, influential signaling is impossible because the incentive for a resolved Deterrer to attempt to surprise the Target is too great. When the Deterrer is highly resolved, it always prefers that the Target not know this. Thus, in this region, the Deterrer never threatens, in the sense that low-cost Deterrer types have an incentive to imitate high-cost types.[50] Once again, diplomatically avoidable wars can occur in equilibrium.

The existence of a communication mechanism lowers the probability of war in general. In the parameterization discussed here, when none exists, the probability of war is 14% for all values of $\overline{\eta}_t$ shown in Figure 6. The reason is that when the Deterrer cannot affect the Target's beliefs, the Target's optimal strategy for all values of η_t is to take the action in question without preparing. Thus, war occurs if and only if the Deterrer prefers to fight one, given that the Target has taken the action and not prepared.

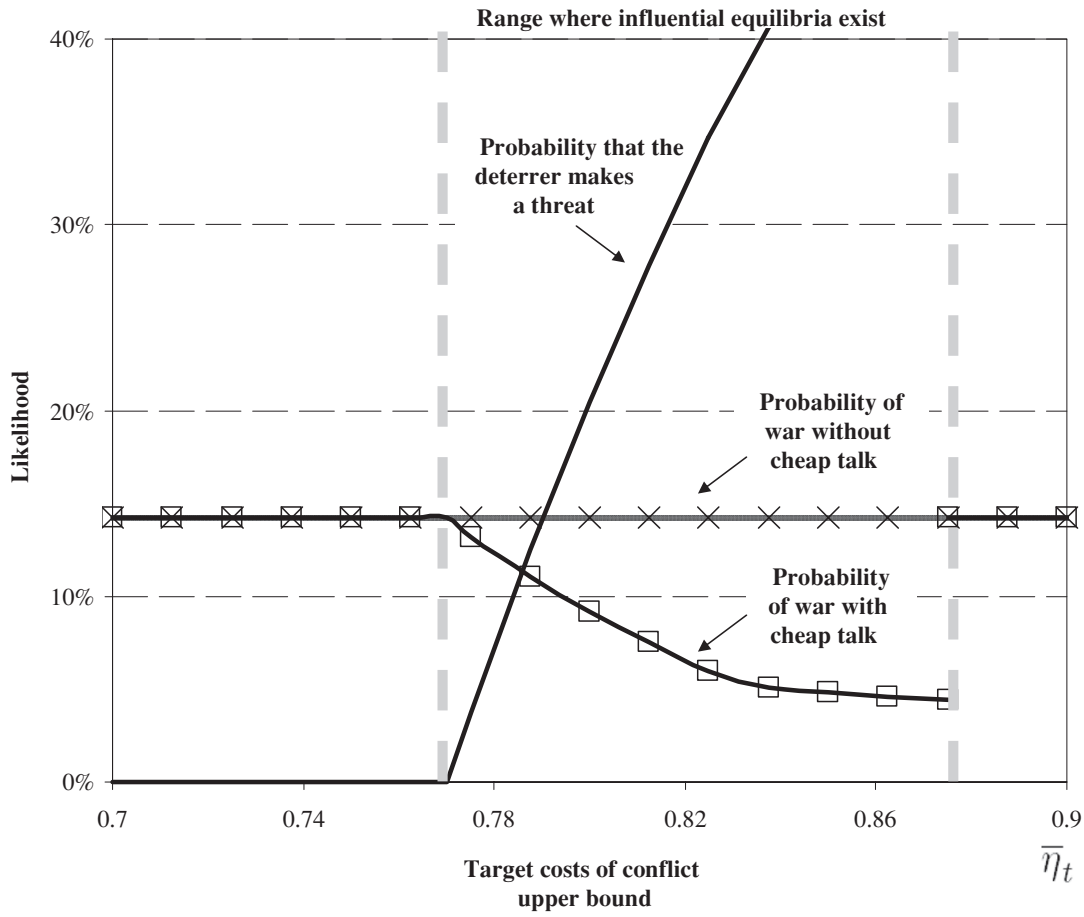
When a communication mechanism exists, the probability of war is reduced over parameter ranges where influential signaling is possible, as Figure 6 illustrates. As we approach the left border of the range where influential equilibria exist from the right, the probability of war is arbitrarily close to the probability of war when no communication mechanism exists. This is so for two reasons. First, even though threats are very informative, as shown on the left-hand side of Figure 5, they are rarely used, as can be seen from the upward-sloping line in Figure 6. When the Deterrer is resolved, it often prefers the gamble of a surprise attack to the gamble of diplomatic signaling. Second, when $\overline{\eta}_t$ is relatively low, there is a low probability that the Target is willing to back down. As $\overline{\eta}_t$ increases, so that the Target is more likely to be a high-cost type, even though threats are less informative, they become more used, surprise attacks decline, and the Target is more likely to be willing to back down, with the result that the probability of conflict decreases. When

[48] See, in particular, Fearon (1995).

[49] Fearon (1995, 395–6) also mentions this incentive to conceal information: "States can also have an incentive to conceal their capabilities or resolve if they are concerned that revelation would make them militarily (and hence politically) vulnerable or would reduce the chances for a successful first strike."

[50] In a cheap talk model, messages have no inherent meaning, so if there is an equilibrium in which all types coordinate on $m = 0$, there is also an equilibrium in which all types send $m = 1$. Nevertheless, we say that the Deterrer "never threatens" in the left-hand range of Figures 4 and 5 because the resolved types have an incentive to mimic the unresolved types.

FIGURE 6. Probability of War



$\overline{\eta}_t = 0.875$, for instance, the probability of war is 4%, less than one-third of what it would be in the absence of a communication mechanism.

When $\overline{\eta}_t$ increases enough so that signaling becomes impossible, however, the probability of war jumps discontinuously back from 4% to 14% for the reason given above. If we were to consider parameterizations where $\overline{\eta}_t$ was much higher, increases in $\overline{\eta}_t$ would again cause the probability of war to decline because Targets for whom war was extremely costly would either decline to take the action in question or make preparations such that the Deterrer would itself be deterred from initiating a conflict. Thus, communication results in a nonmonotonic relationship between the probability that the sides are willing to fight and the probability of war.

CONCLUSION

In international politics, the anarchic strategic context is such that states must develop their foreign policies under conditions of uncertainty about the foreign policies of others. The structure of the system leaves a wide scope for agency in some cases, if not in all, and different leaders react differently to the same foreign policy questions. States also tend to adopt more adversarial policies against states they believe are more threatening to their interests. States have a wide range of options in this regard, such as increasing arms production, forming hostile alliances, first strikes, draining resources, and joining the opposing side in conflicts. Taken together, these aspects of the system imply that costless signals of resolve sent by adversaries can convey information. Diplomatic communication shapes states' perceptions of the threat they pose to each other. A second result of conceiving of the Target of a threat as having a wider array of responses is that conflict can occur because resolved states do not communicate their private information even when they could. In contrast to most other crisis bargaining models, the resolved sometimes have incentive to imitate the behavior of the unresolved.

For the sake of clarity, the model presented in the previous section considered a particular crisis in isolation, but some state decisions have far-reaching effects, and this has implications for the dynamics described here. A key aspect of the model is the possibility that threat-making can result in outcomes that are worse, from the point of view of the threatening state, than allowing the Target to take the action in question. In the model of a particular crisis, this results from the possibility that a Target might elect to prepare for

war, and having prepared, to fight one. More generally, however, the reaction of a Target may have a negative effect on the Deterrer's utility even if no war is fought or contemplated. If the Target builds arms or forms an alliance, its increased capabilities may tip the balance of power and result in diminished bargaining outcomes for the Deterrer down the road. Thus, even in the event of a peaceful outcome in the current crisis, the Target's response to a threat may not be welcomed by the Deterrer, because the latter feels its security decreased as the Target forms hostile alliances, mobilizes, builds arms, or initiates other policies to drain the Deterrer's resources.

Similarly, the incentive of a threatened Target to decline to make preparations when it declines to resist the Deterrer in a particular crisis may not hold in a more general context. If the Target learns from a threat that the Deterrer defines its interests as more opposed to the Target's than the Target had previously thought, the Target may elect to form an alliance with a third state that would be detrimental to the interests of the Deterrer, even when the Target has no plans to contest the particular issue of the current crisis. In these international contexts, the dynamics of signaling are likely to be similar to those in our model of an isolated international crisis; the same essential logic of costless diplomacy will apply.

Whether information can be conveyed in particular cases depends on the context of beliefs and strategic options in which costless statements are made. For instance, there must be a significant *risk*, but not a near certainty, of a breach in relations. Put differently, there must be a possibility that a threatened state will react against the threatener, not merely by refusing to cooperate on the issue at hand, but with respect to other aspects of the relationship. Up to a point, the greater the likelihood of a breach, and the more serious its consequences for the Deterring state, the greater the change in the Target's perceptions of the Deterrer's intentions as a result of a threat. If the likelihood of a breach is too great or its consequences too severe, however, highly resolved Deterrers may have incentive to deceive Targets by refraining from threatening behavior while planning to attack, thereby catching the Target unprepared. Because Targets are aware of Deterrers' incentives, this dynamic can impair the ability of the Deterrer to credibly signal peaceful intentions.

These considerations have several implications for the study of interstate coercion. First, we should not always expect that the quality of the Target's options to prepare for and respond to hostilities will be inversely related to the credibility of the Deterrer's threats. Because, in some contexts, the increased efficacy of Target options will make states more reticent to threaten in the first place, such signals will carry more weight when they are actually used. Second, as in models of public signaling and signaling based on reputation, beliefs are intersubjective. The statements and actions of other states are interpreted in light of what is believed about these states, and also what these states are known to believe about one's own state. If both sides understand that the Deterrer believes the Target fairly likely to back down, for instance, statements by the Deterrer will have little marginal effect on the Target's beliefs about the Deterrer's intentions. Third, the dynamics of communication introduce unexpected nonmonotonicities into the relationship between the factors that affect the likelihood that states are willing to fight over an issue and the probability of war. A decrease in the probability that the Target of a threat is willing to fight can increase the probability of war because of the effect of the Target's perceived likely resolve on the possibility of communication. If the decrease in perceived Target resolve makes influential signaling impossible, as is sometimes the case, this will increase the likelihood of war. Fourth, sometimes resolved states have an incentive to hide their resolve, which results in surprise attacks and wars that could have been avoided if states had just informed each other of their willingness to contest the issues of the day.

The mechanism described here fits many cases. In understanding the dynamics of threat-making, it is instructive to study the fears of decision-makers contemplating such action. Very often, they worry not about the consequences for their reputations of being caught in a bluff, but about the effect on their security positions if their threats are believed. In a multipolar context, Bismarck's tacit threat to Russia in 1876 during the Great Eastern Crisis conveyed information because all sides understood the danger that a frustrated Russia would form an alliance with France. In a bipolar context, Kennedy and Khrushchev's threats to each other over the status of Berlin in June of 1961 conveyed information because the two sides understood the danger that each would increase its efforts to harm the interests of the other side, exacerbating the security dilemma dynamic between the two countries.[51]

The 1876 and 1961 cases are examples of private diplomacy, but we should expect to find the mechanism discussed here operating alongside other mechanisms of information transmission in cases of public diplomacy as well. When U.S. President G. W. Bush says that Iran is part of an "axis of evil," for instance, this may constitute a tacit threat to take actions against the country's leadership.[52] If Iran is unwilling to comply with U.S. demands, it may adopt policies designed to drain U.S. capabilities, for instance, by frustrating U.S. objectives in Iraq. By demonstrating a willingness to risk such a response, the President's statement may convey information about U.S. resolve to force changes in Iranian policy. A similar dynamic is at work when U.S. Defense Secretary Robert Gates says that unless Russia acquiesces to Western demands in Georgia, "the U.S.–Russian relationship could be adversely affected for years to come." If Russia does not comply and therefore believes the U.S. to be more hostile to its interests, Russia will reorient its security posture in ways that have negative consequences for both sides. Though such statements must be interpreted in the light

[51] See *Foreign Relations of the United States*, 1961–1963 (1993), and Jervis (2001).

[52] White House Press Release, January 29, 2002, "President Delivers State of the Union Address."

of a multitude of factors, including subsequent interactions between the two states, by increasing the risk of this negative outcome, these statements convey information. In both public and private contexts, therefore, the existence of costless communication mechanisms is likely to have important effects on conflict processes and the likelihoods of war and peace.

Capabilities alone do not determine outcomes, communication between states is not purely epiphenomenal on clashes of power and interest, and it is not the case, as even the diplomatic historian A. J. P. Taylor has written, that, "wars make the decisions; diplomacy merely records them."[53] There is wide scope for human agency and diplomats in particular in shaping the form that clashing interests will take. Diplomatic conversations, even those between adversaries, are an integral part of the processes that construct perceptions of intention and determine the course of events in the international system.

APPENDIX

A strategy for the Deterrer is a pair $(m(\eta_d), r_d(\eta_d, h))$, where $h \in H \equiv M \times A_1 \times A_2$. A strategy for the Target is a triple $(a_1(\eta_t, m), a_2(\eta_t, m), r_t(\eta_t, h))$. Let the updated beliefs of the Target following the Deterrer's Stage 1 signal be $\mu_{\eta_d}(\eta_d \mid m)$. Let $p(0) = p(1) = p$ if $p(0) = p(1)$. For some particular PBE, let $q_d(a_2, m) = \Pr(r_d = 0 \mid a_1 = 1, a_2, m)$, which is induced by Φ_{η_d} and the Deterrer's strategy, and let $q_t(m)$ be the probability that the Target's strategy satisfies $a_1 = 0, r_t = 0$ for some signal m (which can be written $\Pr(a_1 = 0 \mid m)\Pr(r_t = 0 \mid a_1 = 0, m)$) induced by Φ_{η_t} and the Target's strategy.

The following two lemmas are used in the proof of Proposition 1. Without loss of generality, we suppress the a_2 notation in the proofs of the following lemmas and Proposition 1 so that the Target's strategy is a pair, $(a_1(\eta_t, m), r_t(\eta_t, h))$, and $q_d(a_2, m)$ is instead written $q_d(m)$.

LEMMA 1. *In an influential PBE with ineffective preparations, if* $q_t(m') > q_t(m'')$, $m(\eta_d) = m' \; \forall \eta_d$ *such that* $p - \eta_d > s - \epsilon$.

Proof. Suppose $p - \eta_d > s - \epsilon$. The Deterrer can set $r_d = 1$ and obtain its highest expected utility in any subgame following $a_1 = 1$: $p - \eta_d$. Therefore, any equilibrium strategies must result in an outcome with this Deterrer utility in any subgame following $a_1 = 1$. By assumption, $s > p - \eta_d \; \forall \eta_d$, so optimality of the Deterrer's Stage 3 choice implies $r_d(\eta_d, h) = 0$ for all h such that $a_1 = 0$ when $q_t(m) > 0$ (which implies $\Pr(r_t = 0 \mid a_1 = 0, m) > 0$). Therefore, $\forall \eta_d < p - s + \epsilon$, we can write the Deterrer's expected utility in an equilibrium as a function of its signal: $Eu_d(m) = q_t(m)s + (1 - q_t(m))(p - \eta_d)$, which is increasing in $q_t(m)$. ■

LEMMA 2. *In any influential PBE with ineffective preparations,* $q_d(m) > 0 \; \forall m$.

Proof. We shall use a proof by contradiction: suppose $q_d(\tilde{m}) = 0$ for $\tilde{m} \in M$. Note that in an influential equilibrium, both signals $m = 0$ and $m = 1$ must be sent with positive probability.

First, observe that

$$q_t(\tilde{m}) = 1 - \Phi_{\eta_t}(s - p) > 0. \tag{3}$$

This follows because in any PBE, the expected Target utilities in a subgame following $\tilde{m}$ are as follows: $Eu_t(a_1 = 1, r_t = 0 \mid \tilde{m}) = Eu_t(a_1 = 1, r_t = 1 \mid \tilde{m}) = Eu_t(a_1 = 0, r_t = 1 \mid \tilde{m}) = 1 - p - \eta_t$, $Eu_t(a_1 = 0, r_t = 0 \mid \tilde{m}) = 1 - s$. (The last expected utility uses that, in an influential PBE, $r_d(\eta_d, h) = 0 \; \forall h$ such that $a_1 = 0$. If $\Pr(r_t = 0 \mid a_1 = 0, m = \tilde{m}) > 0$, this follows directly. If $\Pr(r_t = 0 \mid a_1 = 0, m = \tilde{m}) = 0$, then $q_t(\tilde{m}) = 0$. For the equilibrium to be influential, we need $q_t(\check{m}) > 0$ for $\check{m} \neq \tilde{m}$. But then $m(\eta_d) = \check{m} \; \forall \eta_d < p - s + \epsilon$ by Lemma 1, which implies either that $q_d(\tilde{m}) \neq 0$, which contradicts our original assumption, or that $\tilde{m}$ is not sent with positive probability in which case the equilibrium is not influential.) $q_t(\tilde{m})$ is therefore the probability that $1 - s > 1 - p - \eta_t$.

Second, observe that in equilibrium,

$$q_d(\check{m}) = 1 \text{ for } \check{m} \neq \tilde{m} \tag{4}$$

and

$$q_t(\check{m}) = 1 - \Phi_{\eta_t}\left(s - p + \epsilon \frac{q_d(\check{m})}{1 - q_d(\check{m})}\right) < q_t(\tilde{m}). \tag{5}$$

To see this, note that in equilibrium,

$$r_t(\eta_t, h) = 0 \; \forall \eta_t, h \text{ such that } a_1 = 1. \tag{6}$$

Then, let $\Pr(\tilde{m})$ be the probability the Deterrer sends signal $\tilde{m}$ induced by the Deterrer's strategy, so that, in a PBE, $q_d(\check{m}) = \frac{1 - \Phi_{\eta_d}(p - s + \epsilon)}{1 - \Pr(\tilde{m})} > 0$. The Target's expected utilities over strategies in the subgame following $\check{m}$ are the same as its expected utilities following $\tilde{m}$, except that $Eu_t(a_1 = 1, r_t = 0 \mid \check{m}) = q_d(\check{m})(1 - s + \epsilon) + (1 - q_d(\check{m}))(1 - p - \eta_t) > Eu_t(a_1 = 0, r_t = 1 \mid \check{m})$. Thus, in an influential PBE, the Target's strategy in the subgame following signal $\check{m}$ must satisfy the following three conditions: (1) $a_1(\eta_t, \check{m}) = 1 \; \forall \eta_t < s - p + \epsilon \frac{q_d(\check{m})}{1 - q_d(\check{m})}$, (2) $a_1(\eta_t, \check{m}) = 0 \; \forall \eta_t > s - p + \epsilon \frac{q_d(\check{m})}{1 - q_d(\check{m})}$, and (3) $r_t(\eta_t, h) = 0 \; \forall \eta_t > s - p + \epsilon \frac{q_d(\tilde{m})}{1 - q_d(\check{m})}$, h such that $m = \check{m}$. This implies that $q_t(\check{m}) = 1 - \Phi_{\eta_t}(s - p + \epsilon \frac{q_d(\tilde{m})}{1 - q_d(\check{m})}) < q_t(\tilde{m})$. Using Lemma 1, this implies that

$$m(\eta_d) = \tilde{m} \; \forall \eta_d \text{ such that } p - \eta_d > s - \epsilon. \tag{7}$$

Using Bayes' rule, this in turn implies (4).

Third, I show that

$$m(\eta_d) = \tilde{m} \; \forall \eta_d < p - s + \frac{\epsilon}{1 - q_t(\tilde{m})}. \tag{8}$$

Given (4), the Target's optimal strategy in the Stage 2 subgame must satisfy $a_1(\eta_t, \check{m}) = 1 \; \forall \eta_t$. Using (6), this implies that for η_d such that $p - \eta_d < s - \epsilon$, $Eu_d(\check{m} \mid \eta_d) = s - \epsilon$ and $Eu_d(\tilde{m} \mid \eta_d) \geq q_t(\tilde{m})s + (1 - q_t(\tilde{m}))(p - \eta_d)$. Using (7), this implies (8).

Finally, observe that because $q_t(\tilde{m}) > 0$, for h such that $m = \tilde{m}$,

$$r_d(\eta_d, h) = 0 \; \forall \eta_d > p - s + \epsilon. \tag{9}$$

Thus, since Φ_{η_d} is strictly increasing and $q_t(\tilde{m}) > 0$, in a PBE, Bayes' rule, (8), and (9) do not imply $q_d(\tilde{m}) = 0$, which contradicts our original assumption. ■

Proof of Proposition 1. Suppose $q_t(m') > q_t(m'')$. By Lemma 1, in an influential equilibrium, $m(\eta_d) = m' \; \forall \eta_d < p - s + \epsilon$. We have already observed that in a PBE $r_t(\eta_t, h) = 0 \; \forall \eta_t, h$ such that $a_1 = 1$. Then, using Lemma 2, the Target's expected utilities in the Stage 2 subgame for strategies and beliefs consistent with a PBE are $Eu_t(a_1 = 0, r_t = 1 \mid m) = 1 - p - \eta_t < Eu_t(a_1 = 1,$

[53] Taylor (1954, 246).

$r_t = 0 \mid m) = q_d(m)(1-s+\epsilon) + (1-q_d(m))(1-p-\eta_t)$. Thus, in any PBE, if $r_t(\eta_t, h) = 1$ for h such that $a_1 = 0$, then $a_1(\eta_t, m) = 1\ \forall \eta_t, m$. This implies that in a PBE, for η_d such that $p - \eta_d \leq s - \epsilon$,

$$Eu_d(m) = q_t(m)s + (1 - q_t(m))(s - \epsilon).$$

Because this is increasing in $q_t(m)$, in any PBE, $m(\eta_d) = m'\ \forall \eta_d$ such that $p - \eta_d \leq s - \epsilon$. Since m'' is not sent with positive probability in equilibrium, no PBE can be influential. ■

Proof of Proposition 2. Let the Deterrer's strategy be $m(\eta_d) = 1\ \forall \eta_d < \hat{\eta}_d$ and $m(\eta_d) = 0\ \forall \eta_d \geq \hat{\eta}_d$; $r_d(\eta_d, h) = 1$ iff $\eta_d < p(a_2) - s + \epsilon$ & h such that $a_1 = 1$. Let the Stage 3 component of the Target's strategy be as follows: for h such that $a_1 = 1$ & $a_2 = 1$, $r_t(\eta_t, h) = 1$ iff $\eta_t < s - p(1) - \epsilon + \beta_t \equiv \check{\eta}_t$; for h such that $a_1 = 1$ & $a_2 = 0$, $r_t(\eta_t, h) = 0\ \forall \eta_t$; for h such that $a_1 = 0$, $r_t(\eta_t, h) = 1$ iff $1 - p(a_2) - \eta_t + \beta_t a_2 > 1 - s$. Let the Stage 2 component of the Target's strategy be $a_1(\eta_t, 0) = 1, a_2(\eta_t, 0) = 0\ \forall \eta_t$; $a_1(\eta_t, 1) = 1, a_2(\eta_t, 1) = 1\ \forall \eta_t < \hat{\eta}_t$; $a_1(\eta_t, 1) = 0, a_2(\eta_t, 1) = 0\ \forall \eta_t \geq \hat{\eta}_t$ where $\hat{\eta}_t \equiv s - p(1) - k_t + \beta_t$. Let the Target's updated beliefs be $\mu_{\eta_d}(\eta_d \mid 1) = \frac{\Phi_{\eta_d}(\eta_d)}{\Phi_{\eta_d}(\hat{\eta}_d)}\ \forall \eta_d \in [\underline{\eta}_d, \hat{\eta}_d)$ and 0 otherwise, and $\mu_{\eta_d}(\eta_d \mid 0) = \frac{\Phi_{\eta_d}(\eta_d) - \Phi_{\eta_d}(\hat{\eta}_d)}{1 - \Phi_{\eta_d}(\hat{\eta}_d)}\ \forall \eta_d \in [\hat{\eta}_d, \bar{\eta}_d]$ and 0 otherwise.

An equilibrium with these strategies and Target beliefs is influential when $\hat{\eta}_i \in (\underline{\eta}_i, \bar{\eta}_i)\ \forall i$. Note that the updated beliefs of the Deterrer must also follow from Bayes's rule in the equilibrium, but that the Deterrer's optimal actions in the third stage do not depend on these beliefs. The Target's beliefs above follow directly from Bayes' rule, given the Deterrer's strategy. I now show that for some Φ_{η_d} and Φ_{η_t}, if $k_t \geq \epsilon$, the above strategies are optimal and that $\hat{\eta}_i \in (\underline{\eta}_i, \bar{\eta}_i)\ \forall i$.

The optimality of the Stage 3 component of the Target's strategy follows directly from the Target's preferences over outcomes. Given the Target's strategy, the optimality of the Stage 3 component of the Deterrer's strategy follows by backwards induction.

We turn now to the optimality of the Deterrer's Stage 1 choice. Given the Stage 2 and 3 components of the players' strategies, because $\hat{\eta}_t \leq \check{\eta}_t$ because $k_t \geq \epsilon$, the Deterrer's expected utility from threatening is

$$Eu_d(m = 1 \mid \eta_d) = \Phi_{\eta_t}(\hat{\eta}_t)(p(1) - \eta_d) + (1 - \Phi_{\eta_t}(\hat{\eta}_t))s, \quad \textbf{(10)}$$

which is decreasing in η_d because our assumptions imply that $\hat{\eta}_t \in (\underline{\eta}_t, \bar{\eta}_t)$. Given the Stage 2 and 3 components of the players' strategies, the Deterrer's expected utility from not threatening is

$$Eu_d(m = 0 \mid \eta_d) = \begin{cases} p(0) - \eta_d & \eta_d < \tilde{\eta}_d \\ s - \epsilon & \eta_d \geq \tilde{\eta}_d, \end{cases}$$

where $\tilde{\eta}_d = p(0) - s + \epsilon > \underline{\eta}_d$.

Type $\eta_d = \underline{\eta}_d$ prefers to threaten when

$$\Phi_{\eta_t}(\hat{\eta}_t) \leq \frac{s - p(0) + \underline{\eta}_d}{s - p(1) + \underline{\eta}_d} \equiv \ell_t \in (0, 1). \quad \textbf{(11)}$$

We can choose Φ_{η_t} such that condition (11) holds. For instance, let Φ_{η_t} be such that $\Phi_{\eta_t}(\hat{\eta}_t) = \ell_t$. Then

$$Eu_d(m = 0 \mid \eta_d = \underline{\eta}_d) = Eu_d(m = 1 \mid \eta_d = \underline{\eta}_d). \quad \textbf{(12)}$$

Note also that

$$\frac{\partial Eu_d(m = 0 \mid \eta_d < \tilde{\eta}_d)}{\partial \eta_d} < \frac{\partial Eu_d(m = 1 \mid \eta_d < \tilde{\eta}_d)}{\partial \eta_d}. \quad \textbf{(13)}$$

By (12) and (13), note that if $Eu_d(m = 0 \mid \hat{\eta}_d) = Eu_d(m = 1 \mid \hat{\eta}_d)$ for $\hat{\eta}_d \neq \underline{\eta}_d$, then $\hat{\eta}_d > \tilde{\eta}_d$, and in particular,

$$\hat{\eta}_d = p(1) - s + \epsilon / \Phi_{\eta_t}(\hat{\eta}_t).$$

Below, we shall choose $\bar{\eta}_d$ such that $\bar{\eta}_d > \hat{\eta}_d$. With this specification of $\hat{\eta}_d$ and Φ_{η_t}, and the expected utilities given above, we can easily check that the Deterrer's Stage 1 strategy is optimal given the players' beliefs and the Stage 2 and 3 components of the players' strategies. Using (12) and (13) again, we see that $Eu_d(m = 1 \mid \eta_d) \geq Eu_d(m = 0 \mid \eta_d)\ \forall \eta_d < \hat{\eta}_d$. Because $\frac{\partial Eu_d(m=0 \mid \eta_d \geq \tilde{\eta}_d)}{\partial \eta_d} = 0$ and $\Phi_{\eta_t}(\hat{\eta}_t) > 0$, $\frac{\partial Eu_d(m=1 \mid \eta_d)}{\partial \eta_d} < 0$ and $Eu_d(m = 0 \mid \eta_d) \geq Eu_d(m = 1 \mid \eta_d)\ \forall \eta_d > \hat{\eta}_d$. Thus, using the one-stage deviation property, the Deterrer's strategy is optimal given the Target's strategy and the players' beliefs.

We turn now to the optimality of the Target's strategy in Stage 2 subgames. Given the Deterrer's strategy and our definition of $\hat{\eta}_t$, the Target's equilibrium strategy in Stage 3, and the Target's beliefs, the Stage 2 component of the Target's strategy at the $m = 0$ node $(a_1(\eta_t, 0) = 1, a_2(\eta_t, 0) = 0)\ \forall \eta_t$, gives the Target its best outcome $(1 - s + \epsilon)$ with certainty, so the Target cannot gain by deviating.

To complete the proof, it remains only to show that the Target's strategy is optimal in $m = 1$ subgames. Recall that $\check{\eta}_t \equiv s - p(1) + \beta_t - \epsilon$, and let $\dot{\eta}_t \equiv s - p(1) + \beta_t$ and $\ddot{\eta}_t \equiv s - p(0)$. Then, given the Stage 3 component of the players' strategies and the Target's updated beliefs at $m = 1$ nodes, we can write the Target's expected utilities over its four actions at Stage 2 nodes as follows:

$$Eu_t(a_1 = 1, a_2 = 1 \mid m = 1, \eta_t)$$

$$= \begin{cases} 1 - p(1) - \eta_t - k_t + \beta_t & \eta_t < \check{\eta}_t \\ \frac{\Phi_{\eta_d}(p(1)+\epsilon-s)}{\Phi_{\eta_d}(\hat{\eta}_d)}(1 - p(1) - \eta_t - k_t + \beta_t) & \\ \quad + (1 - \frac{\Phi_{\eta_d}(p(1)+\epsilon-s)}{\Phi_{\eta_d}(\hat{\eta}_d)})(1 - s + \epsilon - k_t) & \eta_t \geq \check{\eta}_t \end{cases}$$

(note that $p(1) + \epsilon - s \leq \hat{\eta}_d$);

$$Eu_t(a_1 = 1, a_2 = 0 \mid m = 1, \eta_t)$$

$$= \min\left(1, \frac{\Phi_{\eta_d}(p(0) + \epsilon - s)}{\Phi_{\eta_d}(\hat{\eta}_d)}\right)(1 - p(0) - \eta_t)$$

$$+ \left(1 - \min\left(1, \frac{\Phi_{\eta_d}(p(0) + \epsilon - s)}{\Phi_{\eta_d}(\hat{\eta}_d)}\right)\right)(1 - s + \epsilon)$$

(note that $\frac{\Phi_{\eta_d}(p(0)+\epsilon-s)}{\Phi_{\eta_d}(\hat{\eta}_d)} > 0$ because $s - \epsilon < p(0) - \underline{\eta}_d$ by assumption);

$$Eu_t(a_1 = 0, a_2 = 1 \mid m = 1, \eta_t)$$

$$= \begin{cases} 1 - p(1) - \eta_t - k_t + \beta_t & \eta_t < \dot{\eta}_t \\ 1 - s - k_t & \eta_t \geq \dot{\eta}_t \end{cases};$$

$$Eu_t(a_1 = 0, a_2 = 0 \mid m = 1, \eta_t)$$

$$= \begin{cases} 1 - p(0) - \eta_t & \eta_t < \ddot{\eta}_t \\ 1 - s & \eta_t \geq \ddot{\eta}_t \end{cases}.$$

Because $k_t \geq \epsilon$ and $[(1 - p(1)) - (1 - p(0))] > k_t - \beta_t$, $\dot{\eta}_t > \check{\eta}_t \geq \hat{\eta}_t > \ddot{\eta}_t$. We first show that Target types $\eta_t < \hat{\eta}_t$ have no incentive deviate from the proposed equilibrium. For $\eta_t < \hat{\eta}_t$, the Target has no incentive to deviate to $(a_1(\eta_t, 1) = 0, a_2(\eta_t, 1) = 1)$ because $Eu_t(a_1 = 0, a_2 = 1 \mid m = 1, \eta_t) = Eu_t(a_1 = 1, a_2 = 1 \mid m = 1, \eta_t)\ \forall \eta_t < \hat{\eta}_t$. Since

$Eu_t(a_1 = 1, a_2 = 1 \mid m = 1, \eta_t) > Eu_t(a_1 = 0, a_2 = 0 \mid m = 1, \eta_t)\ \forall \eta_t < \ddot{\eta}_t$ and $Eu_t(a_1 = 1, a_2 = 1 \mid m = 1, \eta_t) \geq Eu_t(a_1 = 0, a_2 = 0 \mid m = 1, \eta_t)\ \forall \eta_t \in [\dot{\eta}_t, \hat{\eta}_t]$, Target types $\eta_t < \hat{\eta}_t$ have no incentive to deviate to $(a_1(\eta_t, 1) = 0, a_2(\eta_t, 1) = 0)$. Further, $Eu_t(a_1 = 1, a_2 = 1 \mid m = 1, \eta_t) \geq Eu_t(a_1 = 1, a_2 = 0 \mid m = 1, \eta_t)\ \forall \eta_t \in [\underline{\eta}_t, \hat{\eta}_t)$ when

$$\frac{\Phi_{\eta_d}(p(0) + \epsilon - s)}{\Phi_{\eta_d}(\hat{\eta}_d)} \geq \frac{-s + p(1) + \epsilon + \eta_t + k_t - \beta_t}{-s + p(0) + \epsilon + \eta_t}.$$

Because the RHS is increasing in η_t, let

$$\ell_d \equiv \frac{-s + p(1) + \epsilon + \overline{\eta}_t + k_t - \beta_t}{-s + p(0) + \epsilon + \overline{\eta}_t},$$

so that, taking the other components of the players' strategies as given, $Eu_t(a_1 = 1, a_2 = 1 \mid m = 1, \eta_t) > Eu_t(a_1 = 0, a_2 = 1 \mid m = 1, \eta_t)\ \forall \eta_t \in (\underline{\eta}_t, \hat{\eta}_t)$ if $\frac{\Phi_{\eta_d}(p(0)+\epsilon-s)}{\Phi_{\eta_d}(\hat{\eta}_d)} > \ell_d$. Below, we shall choose Φ_{η_d} so that this is true. (Note that our assumptions, in particular (2) and $(1 - p(1)) - (1 - p(0)) > k_t - \beta_t$, imply that $\ell_d \in (0, 1)$.)

We now show that all Target types $\eta_t \geq \hat{\eta}_t$ prefer the action prescribed by the Target's equilibrium strategy in a Stage 2 subgame following $m = 1$ (again taking other elements of the equilibrium as given). $Eu_t(a_1 = 0, a_2 = 0 \mid m = 1, \eta_t = \hat{\eta}_t) \geq Eu_t(a_1 = 1, a_2 = 0 \mid m = 1, \eta_t = \hat{\eta}_t) \Leftrightarrow \frac{\Phi_{\eta_d}(p(0)+\epsilon-s)}{\Phi_{\eta_d}(\hat{\eta}_d)} \geq \frac{\epsilon}{p(0)-p(1)+\beta_t-k_t+\epsilon} \equiv \ddot{\ell}_d \in (0, 1)$. Because $\frac{\partial Eu_t(a_1=1,a_2=0|m=1,\eta_t)}{\partial \eta_t} < \frac{\partial Eu_t(a_1=0,a_2=0|m=1,\eta_t)}{\partial \eta_t}\ \forall \eta_t > \hat{\eta}_t$, if $\frac{\Phi_{\eta_d}(p(0)+\epsilon-s)}{\Phi_{\eta_d}(\hat{\eta}_d)} \geq \ddot{\ell}_d$, $Eu_t(a_1 = 0, a_2 = 0 \mid m = 1, \eta_t) \geq EU_t(a_1 = 1, a_2 = 0 \mid m = 1, \eta_t)\ \forall \eta_t \geq \hat{\eta}_t$. Let $\hat{\ell}_d = \max(\ell_d, \ddot{\ell}_d)$.

We can now choose Φ_{η_d} such that $\overline{\eta}_d > \hat{\eta}_d$ and $\frac{\Phi_{\eta_d}(p(0)+\epsilon-s)}{\Phi_{\eta_d}(\hat{\eta}_d)} = \hat{\ell}_d$.[54] Given such a Φ_{η_d}, $Eu_t(a_1 = 1, a_2 = 1 \mid m = 1, \eta_t)$ is at least as high as the expected utility of any other Stage 2 action $\forall \eta_t < \hat{\eta}_t$. Further, $Eu_t(a_1 = 0, a_2 = 0 \mid m = 1, \eta_t) \geq Eu_t(a_1 = 1, a_2 = 0 \mid m = 1, \eta_t)\ \forall \eta_t > \hat{\eta}_t$ since $\frac{\Phi_{\eta_d}(p(0)+\epsilon-s)}{\Phi_{\eta_d}(\hat{\eta}_d)} \geq \ddot{\ell}_d$.

Because $\ddot{\eta}_t < \hat{\eta}_t \leq \check{\eta}_t < \dot{\eta}_t$, and using the definition of $\hat{\eta}_t$ and the fact that $Eu_t(a_1 = 1, a_2 = 1 \mid m = 1, \eta_t)$ is decreasing in η_t, $Eu_t(a_1 = 0, a_2 = 0 \mid m = 1, \eta_t) \geq Eu_t(a_1 = 0, a_2 = 1 \mid m = 1, \eta_t)\ \forall \eta_t \geq \hat{\eta}_t$ and $Eu_t(a_1 = 0, a_2 = 0 \mid m = 1, \eta_t) \geq Eu_t(a_1 = 1, a_2 = 1 \mid m = 1, \eta_t)\ \forall \eta_t \geq \hat{\eta}_t$. Thus, no Target type can gain by deviating from its equilibrium strategy only in Stage 2. Again using the one stage deviation property, this ensures that the Target's equilibrium strategy is optimal in the Stage 2 subgame, given our specification of Φ_{η_d}. ■

Proof of Proposition 3. We shall first show that the following strategies and beliefs constitute a PBE. Let the Deterrer's strategy be $m(\eta_d) = 0\ \forall \eta_d$, $r_d(\eta_d, h) = 1$ iff $\eta_d < p(a_2) - s + \epsilon$ & h such that $a_1 = 1$, and let the Stage 3 component of the Target's strategy be as follows: for h such that $a_1 = 1$, $r_t(\eta_t, h) = 0$ iff $1 - s + \epsilon - k_t a_2 \geq 1 - p(a_2) - \eta_t - k_t a_2 + \beta_t a_2$; for h such that $a_1 = 0$, $r_t(\eta_t, h) = 1$ iff $1 - p(a_2) - \eta_t + \beta_t a_2 > 1 - s$. The Stage 3 components of the players' strategies follow directly from backward induction. Let the Stage 2 component of the Target's strategy be $a_1(\eta_t, 0) = 1, a_2(\eta_t, 0) = 0\ \forall \eta_t$; $a_1(\eta_t, 1) = 1, a_2(\eta_t, 1) = 1\ \forall \eta_t \leq \tilde{\eta}_t$; $a_1(\eta_t, 1) = 0, a_2(\eta_t, 1) = 0\ \forall \eta_t > \tilde{\eta}_t$. Because the Deterrer pools on $m = 0$, the Target does not update its beliefs following this signal. Off the equilibrium path, let the Target's beliefs be $\mu(\eta_d \mid 1) = 1$ (and note that these beliefs are unconstrained by Bayes's rule in a PBE).

We now verify that the Stage 2 component of the Target's strategy is optimal. Given the other components of the players' strategies and beliefs, the Target's expected utilities from actions in the Stage 2 subgame following $m = 1$ are as follows:

$$Eu_t(a_1 = 1, a_2 = 0 \mid m = 1, \eta_t) = 1 - p(0) - \eta_t,$$

$$Eu_t(a_1 = 0, a_2 = 1 \mid m = 1, \eta_t) = \begin{cases} 1 - s - k_t & \eta_t \geq s - p(1) + \beta_t \\ 1 - p(1) - \eta_t - k_t + \beta_t & \eta_t < s - p(1) + \beta_t, \end{cases}$$

$$Eu_t(a_1 = 0, a_2 = 0 \mid m = 1, \eta_t) = \begin{cases} 1 - s & \eta_t \geq s - p(0) \\ 1 - p(0) - \eta_t & \eta_t < s - p(0), \end{cases}$$

and

$$Eu_t(a_1 = 1, a_2 = 1 \mid m = 1, \eta_t) = \begin{cases} 1 - p(1) - \eta_t - k_t + \beta_t & \eta_t < s - p(1) - \epsilon + \beta_t \\ 1 - p(1) - \eta_t - k_t + \beta_t & \eta_t \geq s - p(1) - \epsilon + \beta_t \\ & \quad \& \ p(1) - \underline{\eta}_d > s - \epsilon \\ 1 - s + \epsilon - k_t & \eta_t \geq s - p(1) - \epsilon + \beta_t \\ & \quad \& \ p(1) - \underline{\eta}_d \leq s - \epsilon. \end{cases} \quad \textbf{(14)}$$

Let case (a) be where $k_t > \epsilon$ or $p(1) - \underline{\eta}_d > s - \epsilon$, and let case (b) be where $k_t \leq \epsilon$ and $p(1) - \underline{\eta}_d \leq s - \epsilon$. Let $\tilde{\eta}_t = s - p(1) - k_t + \beta_t$ in case (a), and $\tilde{\eta}_t = \overline{\eta}_t$ in case (b). By comparing the expected utilities for each range of Target types, we can then easily verify that the Stage 2 component of the Target's strategy in the subgame following $m = 1$ is optimal given the other components of the equilibrium.

Given the players' strategies and the Target's beliefs, the Target's expected utilities from actions in the Stage 2 subgame following $m = 0$ are as follows:

$$Eu_t(a_1 = 1, a_2 = 0 \mid m = 0, \eta_t) = q_d(0, 0)(1 - s + \epsilon) + (1 - q_d(0, 0))(1 - p(0) - \eta_t),$$

$$Eu_t(a_1 = 0, a_2 = 1 \mid m = 0, \eta_t) = \begin{cases} 1 - p(1) - \eta_t - k_t + \beta_t & \eta_t < s - p(1) + \beta_t \\ 1 - s - k_t & \eta_t \geq s - p(1) + \beta_t, \end{cases}$$

$$Eu_t(a_1 = 0, a_2 = 0 \mid m = 0, \eta_t) = \begin{cases} 1 - p(0) - \eta_t & \eta_t < s - p(0) \\ 1 - s & \eta_t \geq s - p(0), \end{cases}$$

[54] Note that such a choice of distribution is sufficient but not necessary for the proposed equilibrium to exist. We need only that (1) there is *some possibility* that the Deterrer has high enough costs so that it would not be willing to make a threat in a particular context, and (2) this probability that the Deterrer has high costs is not so great that the Target never takes the Deterrer's threats seriously and thus never makes preparations for conflict.

and

$$Eu_t(a_1 = 1, a_2 = 1 \mid m = 0, \eta_t) = \begin{cases} 1 - p(1) - \eta_t - k_t + \beta_t & \eta_t < s - p(1) - \epsilon + \beta_t \\ q_d(1,0)(1 - s + \epsilon - k_t) & \\ \quad + (1 - q_d(1,0))(1 - p(1) & \\ \quad - \eta_t - k_t + \beta_t) & \eta_t \geq s - p(1) - \epsilon + \beta_t. \end{cases} \quad \textbf{(15)}$$

Because $1 - s + \epsilon$ is the Target's most preferred outcome, we can choose Φ_{η_d} so that $q_d(0,0)$ is high, which implies that $Eu_t(a_1 = 1, a_2 = 0 \mid m = 0, \eta_d)$ is higher than the expected utility of any other Stage 2 option for all η_d in the subgame following $m = 0$. To see this explicitly, let $a = \max(1 - p(1) - \underline{\eta}_t - k_t + \beta_t, 1 - s + \epsilon - k_t)$, which is at least as high as the highest utility that any Target type can achieve from a Stage 2 deviation. We can then calculate the $q_d(0,0)$ required to make $Eu_t(a_1 = 1, a_2 = 0 \mid m = 0, \eta_t) \geq a$ for all η_t:

$$q_d(0,0) \geq \frac{a - 1 + p(0) + \overline{\eta}_t}{-s + \epsilon + p(0) + \overline{\eta}_t} \equiv \hat{q}_d(0,0) < 1. \quad \textbf{(16)}$$

We can see that $\hat{q}_d(0,0) < 1$ by comparing the numerator and denominator of $\hat{q}_d(0,0)$ for all possible values of a. Thus, we can let Φ_{η_d} be such that $\Phi_{\eta_d}(p(0) - s + \epsilon) = 1 - \hat{q}_d(0,0)$, which implies that the Target has no incentive to deviate in the $m = 0$ subgame given the Target does not deviate from its equilibrium strategy in its Stage 3 subgame. Thus, by the one-shot deviation property, the Target's strategy is optimal.

We now turn to the optimality of the Deterrer's strategy in the Stage 1 subgame. The Target's strategy implies that for h such that $(a_1 = 0, a_2 = 0)$, we must have $r_t(\eta_t, h) = 0$ $\forall \eta_t \geq s - p(0)$. Let $\eta'_d \equiv p(1) - s + \epsilon$ and $\eta''_d \equiv p(0) - s + \epsilon$. Then, because $\tilde{\eta}_t > s - p(0)$, given the other components of the players' strategies, we have

$$Eu_d(m = 1 \mid \eta_d) = \begin{cases} (1 - \Phi_{\eta_t}(\tilde{\eta}_t))s + \Phi_{\eta_t}(\tilde{\eta}_t)(p(1) - \eta_d) & k_t > \epsilon \\ (1 - \Phi_{\eta_t}(\tilde{\eta}_t))s + \Phi_{\eta_t}(\tilde{\eta}_t)(p(1) - \eta_d) & k_t \leq \epsilon \text{ and } \eta_d < \eta'_d \\ (1 - \Phi_{\eta_t}(\tilde{\eta}_t))s + (\Phi_{\eta_t}(\tilde{\eta}_t) & \\ \quad - \Phi_{\eta_t}(s - p(1) - \epsilon + \beta_t))(s - \epsilon) & \\ \quad + \Phi_{\eta_t}(s - p(1) - \epsilon + \beta_t)(p(1) - \eta_d) & k_t \leq \epsilon \text{ and } \eta_d \geq \eta'_d, \end{cases} \quad \textbf{(17)}$$

and

$$Eu_d(m = 0 \mid \eta_d) = \begin{cases} s - \epsilon & \eta_d \geq \eta''_d \\ p(0) - \eta_d & \eta_d < \eta''_d. \end{cases} \quad \textbf{(18)}$$

Because $\eta''_d > \eta'_d$, $Eu_d(m = 0 \mid \eta_d < \eta'_d) \geq Eu_d(m = 1 \mid \eta_d < \eta'_d)$ when

$$\Phi_{\eta_t}(\tilde{\eta}_t) \geq \frac{s - p(0) + \eta_d}{s - p(1) + \eta_d}.$$

Because the RHS is increasing in η_d, let $\gamma \equiv \frac{s - p(0) + \overline{\eta}_d}{s - p(1) + \overline{\eta}_d}$ so that $\forall \eta_d < \eta'_d$ $Eu_d(m = 0 \mid \eta_d) \geq Eu_d(m = 1 \mid \eta_d)$ if $\Phi_{\eta_t}(\tilde{\eta}_t) \geq \gamma$. $Eu_d(m = 1 \mid \eta'_d \leq \eta_d)$ is at most $(1 - \Phi_{\eta_t}(\tilde{\eta}_t))s + (\Phi_{\eta_t}(\tilde{\eta}_t) - \Phi_{\eta_t}(s - p(1) - \epsilon + \beta_t))(s - \epsilon) + \Phi_{\eta_t}(s - p(1) - \epsilon + \beta_t)(p(1) - \eta_d) < (1 - \Phi_{\eta_t}(s - p(1) - \epsilon + \beta_t))s + \Phi_{\eta_t}(s - p(1) - \epsilon + \beta_t)(p(1) - \eta_d)$. Setting this less than $Eu_d(m = 0 \mid \eta'_d \leq \eta_d < \eta''_d) = p(0) - \eta_d$ yields

$$\Phi_{\eta_t}(s - p(1) - \epsilon + \beta_t) \geq \frac{s - p(0) + \eta_d}{s - p(1) + \eta_d},$$

which must hold when $\Phi_{\eta_t}(s - p(1) - \epsilon + \beta_t) \geq \gamma$. By similar reasoning, $Eu_d(m = 1 \mid \eta_d \geq \eta''_d) \leq Eu_d(m = 0 \mid \eta_d \geq \eta''_d)$ when

$$\Phi_{\eta_t}(s - p(1) - \epsilon + \beta_t) \geq \frac{\epsilon}{s - p(1) + \eta_d}.$$

Because the RHS is decreasing in η_d, let $\gamma' \equiv \frac{\epsilon}{s - p(1) + \eta''_d} \in (0,1)$. Thus, we can choose Φ_{η_t} to be such that $\Phi_{\eta_t}(\min(\tilde{\eta}_t, s - p(1) - \epsilon + \beta_t)) \geq \max(\gamma, \gamma')$, which implies that, fixing the other components of the equilibrium, the Deterrer has no incentive to deviate from its equilibrium strategy in Stage 1. Thus, the strategies and beliefs given above constitute a PBE for the Φ_{η_t} specified.

Finally, note that (1) this equilibrium is not influential because only one signal is sent with positive probability, and (2) there is a strictly positive probability that $\eta_d \in [\eta_d, p(0) - s + \epsilon)$ and that in case (a) $\eta_t \in (\tilde{\eta}_t, \overline{\eta}_t]$ or in case (b) $\eta_t \in [\underline{\eta}_t, \overline{\eta}_t]$. In either case (a) or case (b), if η_d and η_t are in the ranges specified here, $r = 1$, $m = 0$, and if the Deterrer were to deviate to $m = 1$, no war would occur. ■

REFERENCES

Axelrod, R. 1970. *Conflict of Interest: A Theory of Divergent Goals with Applications to Politics*. Chicago, Markham.

Banks, J. S. 1990. "Equilibrium Behavior in Crisis Bargaining Games." *American Journal of Political Science* 34 (3): 599–614.

Bismarck, O. F. v. 1915. *Gedanken und Erinnerungen*. Berlin: J. G. Cotta'sche Buchhandlung Nachfolger.

Bratman, M. E. 1999. *Intentions, Plans, and Practical Reason*. New York: Center for the Study of Language and Inference.

Chen, Y., N. Kartik, and J. Sobel. 2008. "Selecting Cheap-talk Equilibria." *Econometrica* 76 (1): 117–36.

Crawford, V. P., and J. Sobel. 1982. "Strategic Information Transmission." *Econometrica* 50 (6): 1431–51.

Der Derian, J. 1987. *On Diplomacy: A Geneology of Western Estrangement*. New York: Blackwell.

Farrell, J., and R. Gibbons. 1989. "Cheap Talk Can Matter in Bargaining." *Journal of Economic Theory* 48 (June): 221–37.

Farrell, J., and M. Rabin. 1996. "Cheap Talk." *Journal of Economic Perspectives* 10 (3): 103–18.

Fearon, J. D. 1994. "Domestic Political Audiences and the Escalation of International Disputes." *American Political Science Review* 88 (3): 577–92.

Fearon, J. D. 1995. "Rationalist Explanations for War." *International Organization* 49 (3): 379–414.

Fearon, J. D. 1997. "Signaling Foreign Policy Interests: Tying Hands versus Sinking Costs." *Journal of Conflict Resolution* 41 (1): 68–90.

Fearon, J. D. 1998. "Bargaining, Enforcement and International Cooperation." *International Organization* 52 (2): 269–306.

Feis, H. 1950. *The Road to Pearl Harbor*. Princeton, NJ: Princeton University Press.

Finnemore, M., and K. Sikkink. 1998. "International Norm Dynamics and Political Change." *International Organization* 52 (Autumn): 887–917.

Fursenko, A. A., and T. J. Naftali. 1999. *One Hell of a Gamble: Khrushchev, Castro, Kennedy, and the Cuban Missile Crisis, 1958–1964*. London: Pimlico.

George, A. L. 1991. *Forceful Persuasion: Coercive Diplomacy as an Alternative to War*. Washington, DC: United States Institute of Peace Press.

Gilligan, T. W., and K. Krehbiel. 1987. "Collective Decisionmaking and Standing Committees: An Informational Rationale for Restrictive Amendment Procedures." *Journal of Law, Economics and Organization* 3 (2): 345–50.

Guisinger, A., and A. Smith. 2002. "Honest Threats: The Interaction of Reputation and Political Institutions in International Crises." *Journal of Conflict Resolution* 46 (2): 175–200.

Healy, B., and A. Stein. 1973. "The Balance of Power in International History: Theory and Reality." *Journal of Conflict Resolution* 17 (1): 33–61.

Ignatyev, N. 1931. "The Memoirs of Count N. Ignatyev." *Slavonic Review* 10: 386–407; 627–40.
Jervis, R. 1970. *The Logic of Images in International Relations*. New York: Columbia University Press.
Jervis, R. 1976. *Perception and Misperception in International Politics*. Princeton, NJ: Princeton University Press.
Jervis, R. 1984. "Deterrence and Perception." In *Strategy and Nuclear Deterrence*, ed. S. Miller, Princeton, NJ: Princeton University Press, 57–84.
Jervis, R. 2001. "Was the Cold War a Security Dilemma?" *Journal of Cold War Studies* 3 (1): 36–60.
Kurizaki, S. 2007. "Efficient Secrecy: Public versus Private Threats in Crisis Diplomacy." *American Political Science Review* 101 (3): 543–58.
Kydd, A. 1997. "Game theory and the Spiral Model." *World Politics* 49 (3): 371–400.
Kydd, A. 2003. "Which Side Are You on? Bias, Credibility, and Mediation." *American Journal of Political Science* 47 (4): 597–611.
Kydd, A. 2005. *Trust and Mistrust in International Relations*. Princeton, NJ: Princeton University Press.
May, E. R., and P. D. Zelikow. 2002. *The Kennedy Tapes: Inside the White House during the Cuban Missile Crisis*. New York: Norton.
Mearsheimer, J. J. 2001. *The Tragedy of Great Power Politics*. New York: W. W. Norton.
Mercer, J. 1996. *Reputation and International Politics*. New York: Cornell University Press.
Morgan, P. M. 2003. *Deterrence Now*. New York: Cambridge University Press.
Morrow, J. D. 1989. "Capabilities, Uncertainty, and Resolve: A Limited Information Model of Crisis Bargaining." *American Journal of Political Science* 33 (4): 941–72.
Morrow, J. D. 1994. "Alliances, Credibility, and Peacetime Costs." *Journal of Conflict Resolution* 38: 270–97.
Mosse, W. 1958. *The European Powers and the German Question: 1848–71*. London: Cambridge University Press.
O'Neill, B. 1999. *Honor, Symbols, and War*. Ann Arbor, MI: University of Michigan Press.
Powell, R. 1988. "Nuclear Brinkmanship with Two-Sided Incomplete Information." *American Political Science Review* 82 (1): 155–78.
Powell, R. 1990. *Nuclear Deterrence Theory: The Problem of Credibility*. Cambridge, MA: Cambridge University Press.
Powell, R. 1993. "Guns, Butter and Anarchy." *American Political Science Review* 87 (1): 115–32.
Press, D. G. 2005. *Calculating Credibility: How Leaders Assess Military Threats*. New York: Cornell University Press.
Ramsay, K. W. 2004. "Politics at the Water's Edge: Crisis Bargaining and Electoral Competition." *Journal of Conflict Resolution* 48 (4): 459–86.
Rich, N. 1965. *Why the Crimean War? A Cautionary Tale*. Hanover, NH: University Press for New England.
Ritter, J. M. 2004. "Silent Partners and Other Essays on Alliance Politics." Ph.D. diss., Harvard University.
Rupp, G. H. 1941. *A Wavering Friendship: Russia and Austria 1876–1878*. Cambridge, MA: Harvard University Press.
Russett, B. M. 1967. "Pearl Harbor: Deterrence Theory and Decision Theory." *Journal of Peace Research* 4 (2): 89–106.
Saburov, P. A. 1929. *The Saburov Memoirs or Bismarck and Russia*. Cambridge: Cambridge University Press.
Sartori, A. E. 2002. "The Might of the Pen: A Reputational Theory of Communication in International Disputes." *International Organization* 56 (1): 121–49.
Sartori, A. E. 2005. *Deterrence by Diplomacy*. Princeton, NJ: Princeton University Press.
Schelling, T. C. 1966. *Arms and Influence*. New Haven, CT: Yale University Press.
Schelling, T. C. 1980. *The Strategy of Conflict*. Cambridge, MA: Harvard University Press.
Schultz, K. 1998. "Domestic Opposition and Signaling in International Crises." *American Political Science Review* 92 (4): 829–44.
Schultz, K. 2001. *Democracy and Coercive Diplomacy*. Cambridge, UK: Cambridge University Press.
Schweinitz, H. L. v. 1927. *Denkwuerdigeiten des Botschafters General v. Schweinitz*, Vol. 1. Berlin: Reimar Hobbing.
Schweller, R. 1998. *Deadly Imbalances: Tripolarity and Hitler's Strategy of World Conquest*. New York: Columbia University Press.
Schweller, R. L. 1994. "Bandwagoning for Profit: Bringing the Revisionist State back in." *International Security* 19 (1): 72–107.
Slanchev, B. L. 2005. "Military Coercion in Interstate Crises." *American Political Science Review* 99 (4): 533–47.
Slantchev, B. L. n.d. "Feigning Weakness." *International Organization*. Forthcoming.
Smith, A. 1998. "International Crises and Domestic Politics." *American Political Science Review* 92 (3): 623–38.
Taylor, A. J. P. 1954. *The Struggle for Mastery in Europe, 1848–1918*. Oxford: Clarendon Press.
Trager, R. F. 2007. *Diplomatic Calculus in Anarchy: The Construction and Consequences of the Space of Intentions*. Ph.D. diss., Columbia University.
Walt, S. M. 1987. *The Origins of Alliances*. Ithaca, NY: Cornell University Press.
Waltz, K. N. 1979. *Theory of International Politics*. New York: McGraw–Hill.
Waltz, K. N. 2003. "Evaluating Theories." In *Realism and the Balancing of Power*, eds. C. Elman and J. A. Vasquez, Upper Saddle River, NJ: Prentice Hall, 49–57.
Wendt, A. 1999. *Social Theory of International Politics*. Cambridge, UK: Cambridge University Press.