



---

Diplomatic Signaling among Multiple States

Author(s): Robert F. Trager

Source: *The Journal of Politics*, Vol. 77, No. 3 (July 2015), pp. 635-647

Published by: [The University of Chicago Press](#) on behalf of the [Southern Political Science Association](#)

Stable URL: <http://www.jstor.org/stable/10.1086/681259>

Accessed: 25/10/2015 20:54

---

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <http://www.jstor.org/page/info/about/policies/terms.jsp>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



*The University of Chicago Press and Southern Political Science Association are collaborating with JSTOR to digitize, preserve and extend access to The Journal of Politics.*

<http://www.jstor.org>

# Diplomatic Signaling among Multiple States

---

Robert F. Trager, University of California, Los Angeles

Despite the difficulty of communicating with adversaries, scholars have described a variety of signaling mechanisms that relate to bilateral negotiations between states. This article demonstrates that when more than two states are involved, states have additional, costless means of communicating their intentions. In particular, statements of a third party to a dispute on behalf of a “protégé” are credible because of the effect these statements have on the protégé’s conduct. Protégés that are emboldened by support will sometimes be more likely to take actions that risk conflict, causing the third party to be more likely to have to intervene. Thus, commitments to fight on behalf of other states can convey information to potential adversaries. This form of signaling requires that the interests of the third party and protégé be sufficiently aligned and that the third party be powerful enough, but also not too powerful.

Adversaries have difficulty communicating because they have incentives to mislead (Fearon 1995; Schelling 1966). Nevertheless, scholars have described a variety of costly and costless signaling mechanisms in the case of negotiations between two states (e.g., Fearon 1994a; Guisinger and Smith 2002; Jervis 1970; Kurizaki 2007; Kydd 1997; Sartori 2005; Schelling 1966; Schultz 2001; Trager 2010, 2011). These same mechanisms generally apply when negotiations involve additional states whose interests are closely aligned with one or the other of the adversaries. When more than two states are involved, however, the analysis below demonstrates that states have other means of communicating their intentions. In particular, a costless communication mechanism exists among three or more states that is a common feature of international negotiations: statements of a third party to a dispute on behalf of a “protégé” are credible because of the effect these statements can have on the protégé’s conduct. Unlike other signaling mechanisms, this one does not rely on the sunk costs of alliance commitment or arms production, the need to preserve domestic or international reputations, or other factors previously examined in the literature.

The signaling mechanism analyzed below relates to a central concern of the literature on alliance commitments, namely, the problem of “entrapment” (Jervis, Lebow, and

Stein 1985; Goldstein 2000; Snyder 1997; Zartman and Faure 2005). When a third party makes a commitment to a protégé, these commitments may embolden the protégé. When one state is emboldened by support from another state, the emboldened state may be more likely to take actions that risk conflict with its adversary, causing the third party to be more likely to be forced to intervene to support the protégé if the third party is indeed willing to do so. As a result, commitments to fight on behalf of other states can convey information to potential adversaries even when those commitments are made behind closed doors. In fact, states can sometimes infer from the bargaining behavior of an adversary whether a third party has committed to the defense of the adversary. These sorts of commitments only constitute entrapment when the third party would prefer to hold the protégé back from aggressive actions against the adversary, but the third party then finds it is unable to do so. When the third party knowingly undertakes a risk of conflict through emboldenment, however, this is a form of diplomatic signaling, and one that appears to be common in international politics.

Signals of this sort are likely to influence the calculations of adversaries when all sides understand that the protégé is in a bargaining relationship with an adversary and two other conditions hold. First, the emboldening effect of the third

---

Robert F. Trager (rtrager@ucla.edu) is an associate professor of political science at the University of California, Los Angeles, CA 90095.

Data and supporting materials necessary to reproduce the numerical results in the article are available in the *JOP* Dataverse (<https://dataverse.harvard.edu/dataverse/jop>). An online appendix containing supplemental analyses is available at <http://dx.doi.org/10.1086/681259> and is also available here: <http://www.roberttrager.com/>.

The Journal of Politics, volume 77, number 3. Published online April 23, 2015. <http://dx.doi.org/10.1086/681259>

© 2015 by the Southern Political Science Association. All rights reserved. 0022-3816/2015/7703-0004\$10.00

635

party's commitment on its protégé must increase the likelihood of conflict more than the deterrent effect of the commitment on the adversary state decreases the likelihood of conflict. This implies that the third party must be powerful enough, but also not too powerful. In fact, because it makes credible signaling impossible, an increase in the power of the third party can increase the likelihood of war. Second, the interests of the third party and protégé must be sufficiently aligned.<sup>1</sup>

One example of these dynamics that is of direct, current relevance is the three-way relationship between China, Taiwan, and the United States (Benson and Niou 2005; O'Hanlon 2000). Policy makers in Taiwan are very sensitive to the level of US commitment, and US policy makers are usually careful to affirm some commitment to defend Taiwan in the event of attack, and at the same time express support for the "one China" policy. If the United States were to openly support full Taiwanese independence, it would increase the likelihood that Taiwan will take more concrete steps in that direction. These steps might well increase the likelihood that China would attack. Thus, as a result of making a firm commitment to independence, the United States could well find itself forced to choose between the destruction of Taiwan and a military engagement with China. For this reason, if the United States were to make such a commitment, even behind closed doors, Chinese officials would be right to revise upward their appraisal of the level of US commitment to defend the island.

Another example from the diplomacy of the moment concerns the attempts by the United States and Israel to prevent Iran from acquiring a nuclear weapon. Support from the United States has an emboldening effect on Israel, perhaps making near-term conflict more likely. As a result, US statements of support carry increased credibility. As in the Taiwan case, all parties understand that in offering support, the United States is taking the risk of being forced to fight with its ally or decline to do so, and this risk is entirely separate from any "audience costs" associated with backing down that public US commitments may incur (e.g., Fearon 1994a; Tomz 2007; Trager and Vavreck 2011). Such risks that arise in multistate contexts are equally felt when diplomacy is carried on behind closed doors.

### DIPLOMATIC SIGNALING IN MULTISTATE CONTEXTS

The dynamics of signaling among states have been the subject of sustained analysis by international relations scholars,

1. As discussed below, a number of factors determine whether the third party and protégé's interests are aligned, including the probability the protégé would win in a conflict without aid from the third party and the quality of the bargain the protégé could strike without support.

with theorizing most often focusing on bilateral contexts. All of the formal models cited above fall in this category. Qualitative work, too, has most often examined episodes where one state threatens one other state or coalition of states (but see Crawford 2003). Even when scholars have analyzed signaling in multistate contexts, they have often done so by applying lessons drawn from bilateral analysis (Fearon 1994b).

A substantial body of work examines "extended deterrence," when states commit to defend territories outside their own borders. Studies have analyzed the influence of many factors on deterrence success, including the local military balance (e.g., Huth 1988a; Huth and Russett 1984), states' actions in earlier crises (Huth 1988a, 1988b), the scope of coercive demands (Werner 2000), and the geopolitical stakes in the region (Danilovic 2002), and some studies examine several of these factors in combination with others (Russett 1963). The current study draws on this literature, and the findings from the model analyzed below are often consistent with conclusions from this literature, but the analysis also suggests alternative ways of understanding the influence of some factors on deterrence outcomes. For instance, Huth and Russett (1984) and Danilovic (2002) argue that deterrence is successful when the "defender" has a strong interest in the survival and advancement of the protégé. The analysis below also finds that substantial alignment of the interests of the two is necessary for informative signaling. But this does not imply that signaling is unimportant. Rather, uncertainty about whether a third party would join a potential conflict often remains even when interests appear aligned (Gartner and Siverson 1996; Smith 1996), and it is this uncertainty that costless, diplomatic signals can remove when the defender has a sufficient interest in the protégé. Similarly, the analysis provides a new perspective on the debate over the role of military power in successful coercion (Fearon 1994b; Huth and Russett 1984; Karsten, Howell, and Allen 1984; Lebow and Stein 1989; Maoz 1983; Signorino and Tarar 2006). In the costless signaling model, while sufficient power is a necessary factor in coercion, too high a level of power makes signaling impossible and can thereby decrease the probability of coercive success.<sup>2</sup>

Scholars in this literature have recognized that extended deterrence is importantly different from bilateral coercion (Morgan 1983). It is commonly accepted, for instance, that extended deterrent threats are more difficult to make credible than threats to defend the homeland (Schelling 1966, chap. 2). Nevertheless, researchers have not focused on whether the underlying signaling mechanisms involved mir-

2. For related discussion, see Sechser (2010) and Snyder and Diesing (1977).

ror those of bilateral contests. For instance, one model that does examine three actors is Quackenbush (2006), which argues that when two states are in alliance and each would want to defend the other only if the other would defend it, then the more reliable of the alliance partners is the one that the third state is more likely to attack. This study and a few others examine dynamics among three players, but they do not analyze the effects of statements of commitment (Fang, Johnson, and Leeds 2014; Wagner 2004; Yuen 2009; Zagare and Kilgour 2003).<sup>3</sup> Benson (2012) examines a three-player model with the possibility of emboldenment of a protégé, but focuses on when a third party would make probabilistic, conditional, or firm commitments and therefore assumes rather than demonstrates that the third party has such commitment options available to it. Perhaps the most similar model to the one analyzed below is the three-player costless signaling model investigated in Smith 1998b. This model shows how costless signals can influence beliefs, but, consistent with bilateral signaling models, does so through analysis of the reaction of a domestic constituency to public statements of commitment. The literature on intervention in the internal affairs of other states also considers dynamics among three actors, including when state signals can encourage and embolden a substate group, which is related to the analysis below. These works do not connect the emboldenment of the group to the credibility of the state's signals, however (Cetinyan 2002; Grigoryan 2010; Kuperman 2008). No studies in the international politics literature of three or more actors examine the effect of costless statements of commitment that do not derive their credibility from the reactions of domestic constituencies to public actions.<sup>4</sup>

The literature on the credibility of alliance commitments is also closely related to the topic addressed here. Much of the literature on alliances has focused on factors that make allies more or less likely to fight together (e.g., Benson 2011; Snyder 1997), or examines empirically whether alliances influence adversary calculations (e.g., Huth and Russett 1984; Leeds 2003). Studies that explicitly investigate the mechanism through which alliance commitments influence adversary calculations have generally focused on the public aspect of alliance formation. When alliance commitments are modeled in a way that allows them to convey information, this is primarily a result of the domestic or international reputational costs that backing down from the public commitment would entail (Leeds 1999; Morrow 2000). Scholars have also considered the signaling effects of sunk

costs involved in alliance formation (Morrow 1994; Smith 1995, 1998a) and increased fighting ability through the harmonization of military planning (Morrow 1994; Smith 1998a), as well as the effects of regime type (Gaubatz 1996; Leeds 1999). Thus, these studies also view signaling in multistate contexts as essentially similar to bilateral signaling; the same mechanisms are thought to apply in the same ways.

This article examines a signaling mechanism that is available only in the multistate context, and one that does not depend on reputation or the sunk costs of alliance commitment. Information is communicated in the cheap talk equilibrium through costless signals. As such, the model helps to explain inferences drawn from private diplomatic encounters and informal agreements (Trager 2013). The model described below is intentionally very similar to other models of two players described in the literature. This makes it clear that the changed effects of costless communication are the specific consequence of the signaling possibilities that are created when more than two parties are involved.

The analysis here is consistent with a substantial theoretical literature that notes that increasing the number of players can increase the possibilities for communication and thereby increase the set of equilibrium outcomes. Forges (1990), for instance, shows that with four or more players, any correlated equilibrium outcome of a game can also be an equilibrium outcome of the game with (unmediated) communication between the players. Since the set of correlated equilibria is often larger than the set of Nash equilibria, this shows that communication often increases the set of equilibrium outcomes. With fewer players, however, communication increases the set of equilibrium outcomes only in a more restricted set of games. This literature establishes the possible effects of communication without describing mechanisms that it is reasonable to expect would actually be used by diplomats and leaders, however.

This article, by contrast, develops a mechanism that is very simple and appears to track common considerations of decision makers in international politics. The analysis also shows that *even in the highly adversarial context* of international crisis bargaining, increasing the number of actors enables information to be communicated that otherwise could not be, which is surprising (Crawford and Sobel 1982). The availability of mechanisms for costlessly communicating resolve in turn affects expectations about the outcomes of crises.

### A MODEL OF MULTIPARTY NEGOTIATION

Consider a game in which a "Third Party" (d) sends a costless signal  $m \in M$  to a "Protégé" (g) and a "Target" (t) where  $0, 1 \in M$  and  $M$  is large but finite.<sup>5</sup> The set of players

3. Kilgour and Zagare (1994) examine extended deterrence through a two-player model.

4. Walter (2002) and Clare and Danilovic (2010) discuss incentives to build reputation due to multiple strategic adversaries.

5. In some literatures, the Third Party is called the "Defender."

is  $I \equiv \{d, g, t\}$  with generic element  $i$ . After the signal, a standard crisis bargaining game occurs that can be thought of as a much simplified version of the game in Powell (1996a, 1996b). Following any message  $m$ , the Protégé makes a demand  $x \in [0, 1]$  on the Target. If the Target accepts, the risk-neutral payoffs are  $x$  for the Protégé and  $1 - x$  for the Target. If the Target rejects the offer, the Protégé decides whether to fight or not. If the Protégé does not fight, the status quo is maintained, and payoffs for the Protégé and Target are  $q \in [0, 1]$  and  $1 - q$ , respectively.

If the Protégé fights, the Third Party decides whether or not to join the conflict on the side of the Protégé. In a conflict, the Protégé wins with probability  $p \in (0, 1)$  if the Third Party does not join in and probability  $p^a > p$  if the Third Party does (where  $p^a < 1$ ). The side that wins the conflict chooses its most preferred outcome in  $X$ , 1 for the Protégé and 0 for the Target. The Third Party and Protégé have the same preference orderings over outcomes in  $X$ , but possibly different costs of conflict. Thus, if the Third Party declines to join the conflict, expected utilities for the Protégé, Target, and Third Party are  $p - c_g$ ,  $1 - p - c_t$ , and  $p$ , respectively. If the Third Party fights, expected utilities for the players are  $p^a - c_g$ ,  $1 - p^a - c_t$ , and  $p^a - c_d$ .

Assume that the Protégé prefers the status quo to fighting alone and prefers fighting with the Third Party to the status quo. Formally, this means  $p - c_g < q < p^a - c_g$ . Assume that the Third Party and Target's costs of war are the private information of the player and can be either high or low. Formally, for  $i = d, t$ ,  $c_i = \underline{c}_i$  with probability  $h_i$  and  $c_i = \bar{c}_i$  with probability  $1 - h_i$ , where  $\underline{c}_i < \bar{c}_i$  and  $h_d$  and  $h_t$  are independent. The Third Party prefers to fight with the Protégé if and only if the Third Party has low costs of conflict, so  $p^a - \bar{c}_d < p < p^a - \underline{c}_d$ . Let  $\mu_i(m)$  be player  $i$ 's updated belief that  $c_d = \underline{c}_d$  following message  $m$ .

In two-player contexts of this sort, no information can be conveyed by the Third Party's statements. In the three-player context, however, that is not so. I will show that a fully informative equilibrium can exist in which the two types of Third Party send different messages, so that, upon receiving the message, both the Protégé and Target know for certain whether the Third Party is willing to fight on behalf of the Protégé. In this equilibrium, the Third Party's statements change the Protégé's actions, and this influence on the Protégé's conduct changes the incentives of the Third Party and implies that the Third Party may have no reason to mislead the Target. In game theoretic terms, the equilibrium is fully separating: the Third Party sends a statement of support for the Protégé if and only if the Third Party is willing to fight on the side of the Protégé if necessary; the Protégé makes a high demand that risks conflict if and only if it is supported; the

Target rejects an unsupported Protégé's demand if it is too high (which does not occur in equilibrium) or a supported Protégé's demand if the Target is a low cost type; and, on the equilibrium path, the Protégé fights if and only if it is supported and its high demand is rejected by the Target. For this sort of signaling to be possible at all, several formal conditions must hold. I will develop each of these conditions and the intuition behind them before stating sufficient conditions for the existence of such an equilibrium in a proposition.

First, note that in a signaling equilibrium of this sort, the Protégé and Target must be in a bargaining relationship, the Protégé's optimal behavior must be affected by its expectations about whether the Third Party will offer support in a conflict, and the optimal behavior of a Protégé that expects support must imply a greater probability of war than the optimal behavior of a Protégé that does not expect support. This implies that the Third Party cannot be so powerful that even a more resolute Target would prefer to give up the whole of the issues in dispute rather than wage a war against both the Protégé and Third Party. If the Third Party were so powerful, the Target would be completely deterred from contesting the issues, and the probability of conflict when the support of the Third Party is expected would be zero. This would give the Third Party too large an incentive to offer its support, even if it were not willing to really give it, and signaling would be impossible. Formally, since the expected value of war for the more resolute Target when the Third Party fights is  $1 - p^a - \underline{c}_t$ , signaling is only possible when  $1 - p^a - \underline{c}_t > 0$  or  $p^a + \underline{c}_t < 1$ . Thus, while the Third Party must be strong enough that the Protégé would want to fight rather accept the status quo when the Protégé is assured of Third Party support ( $p^a - c_g > q$  as assumed), this analysis shows that when the Third Party is too powerful ( $p^a$  too high), nothing the Third Party says will affect the beliefs of the other players about what the Third Party is actually willing to do.

For the expectation of support from the Third Party to generate a greater risk of conflict, the Protégé's demand must be large enough that there is a possibility the Target may reject the demand. If this is not so, for any demand that can be made in equilibrium, the probability of conflict is zero and so cannot be larger than in cases where the Protégé does not have an expectation of support. In equilibrium, when it expects support, the Protégé must make either a high demand ( $p^a + \bar{c}_t$ ) or a low demand ( $p^a + \underline{c}_t$ ). If the Protégé makes the low demand, the risk of conflict is zero if the players expect Third Party support in a conflict. Thus, in a signaling equilibrium, the Protégé must be willing to make the high demand, which it is when its expected utility of making that demand ( $h_t(p^a - c_g) + (1 - h_t)(\min\{1, p^a + \bar{c}_t\})$ ) is greater

than its expected utility for the lesser demand ( $p^a + \underline{c}_t$ ). When the Target is willing to fight unless it receives at least some of the good in question so that  $1 - p^a - \bar{c}_t > 0 \Leftrightarrow p^a + \bar{c}_t < 1$ , the Protégé is willing to make the high demand when:

$$h_t \leq \frac{\bar{c}_t - \underline{c}_t}{\bar{c}_t + c_g} \tag{1}$$

In other words, for signaling to occur, it cannot be too certain that the Target is a high-resolve type because if it were, the Protégé would not be willing to make a high demand, and an expectation of Third Party support would merely allow the Protégé to get more without incurring any increased risk of conflict. This, in turn, would make expressing support too tempting for the Third Party to resist, even when it would not be willing to follow through, and this makes signaling impossible.

For costless signaling to occur, the Third Party must of course be willing to reveal its type. If the Third Party reveals that it is not willing to support the Protégé, then no demand for an improvement in the status quo is credible. The Target will understand that the Protégé prefers the status quo to conflict, and the Target will therefore reject any demand greater than  $q$  and the status quo will be the outcome of the game. The first condition for incentive compatibility in a separating equilibrium is therefore that the unresolved Third Party prefer to reveal its type, yielding a payoff of  $q$ , to pretending to be resolved, which yields an expected payoff of  $h_t p + (1 - h_t)(p^a + \bar{c}_t)$  since, given that the Protégé makes a high demand, there is an  $h_t$  chance that the Target rejects the offer and a war is fought in which the Third Party does not join in and a  $(1 - h_t)$  chance that the Protégé's high demand is accepted. Thus, the first incentive compatibility condition is satisfied when:

$$h_t \geq \frac{p^a + \bar{c}_t - q}{p^a + \bar{c}_t - p} \tag{2}$$

Analysis of equation (2) immediately tells us that for signaling to occur, it must be the case that  $p < q$ . If this is not the case, then there is no drawback for the Third Party to pretend to be willing to offer support if it is not willing to do so. In effect, even if the Third Party isn't willing to join in the conflict, it still prefers that the Protégé and Target fight a war to maintaining the status quo. This is so because the Protégé's prospects in the war are sufficiently favorable relative to the current state of affairs. Thus, signaling requires that the Protégé be sufficiently weak. Equation (2) also indicates that there must be a sufficiently high likelihood that the Target is the more resolved type. If this probability is not sufficiently high,

then unresolved Third Parties will again be too tempted to misrepresent their willingness to support their Protégé.

The final incentive compatibility condition for informative signaling is that a resolved Third Party prefer to reveal its type. If it does not do so, as we have seen, a high offer from the Protégé will not be accepted, the Protégé will not elect to fight, and the payoff to the Third Party will be  $q$ . When a resolved Third Party signals its willingness to fight on behalf of the Protégé, the Third Party's expected utility is  $h_t(p^a - \underline{c}_d) + (1 - h_t)(p^a + \bar{c}_t)$ . Thus, resolved Third Parties are willing to reveal their type when:

$$h_t \leq \frac{p^a + \bar{c}_t - q}{\bar{c}_t + \underline{c}_d} \tag{3}$$

For these conditions to be satisfied simultaneously, the parameters must be such that the right-hand side of equations (1) and (3) is greater than the right-hand side of equation (2). Proposition 1 gives sufficient conditions for this to be the case so that a range of values of  $h_t$  produce a separating equilibrium in which the costless signals of the Third Party convey its type. In the proposition, the condition that  $p$  is sufficiently low helps to ensure that less resolved Third Parties will not pretend to be resolved and thereby risk that their relatively weak Protégé finds itself in an unsupported conflict. Low  $\underline{c}_d$  implies that a more resolved Third Party is willing to incur the risk of war that supporting the Protégé entails. The condition on the resolved Target's costs of war,  $\underline{c}_t$ , results from the influence these costs have on the quality of the bargain the Protégé can strike without risking war. The less the Target wants to go to war, the better the bargain for the Protégé. Thus, low  $\underline{c}_t$  implies that the bargain the Protégé can negotiate without risking war may be poor enough that the Protégé decides instead to make an offer that risks conflict when the Protégé's costs of conflict,  $c_g$ , are also sufficiently low.<sup>6</sup>

**Proposition 1.** For  $p, \underline{c}_d, \underline{c}_t, c_g$  sufficiently low,  $p^a + \bar{c}_t < 1$  and  $h_t$  in a middle range, a separating perfect Bayesian equilibrium exists in which  $\mu_i(0) = 0$  and  $\mu_i(1) = 1 \forall i$ .

When the conditions given in proposition 1 are satisfied, the likelihood that the Target is highly resolved is neither

6. Note that since the Third Party's statements are costless, as in all such models, a babbling equilibrium exists in which none of the parties try to communicate and thus none listen either. These equilibria do not appear to correspond to the understandings and intentions of diplomats and state leaders, and I shall not focus on such equilibria.

too great nor too small (see the appendix for a proof of proposition 1). It is not so large that the Third Party would not be willing to risk war in offering to support the Protégé and the Protégé would not be willing to risk conflict in negotiations when it receives such an offer of support. Yet, the likelihood that the Target is highly resolved is not so small that the Third Party would always be willing to incur the risks of conflict that supporting the Protégé entails. When the actors' incentives are in balance in this way, signaling is possible.

To summarize, in the signaling equilibrium, the Protégé makes a high demand if and only if it receives support from the Third Party. The Target accedes to the Protégé's high demand if and only if the Target is not highly resolved. If the Protégé makes a low demand, the Target is certain to agree to that as well. If the Target does not accept the Protégé's demand and the Protégé has support from the Third Party, the Protégé will go to war. Thus, support from the Third Party improves the Protégé's bargaining position, but also increases the probability of war. Nevertheless, since only resolved Third Parties are willing to embroil their Protégés in a potential conflict, the Third Party's signal conveys information.

Several important extensions to the model are considered in the online supplementary appendix to the article. These show that signaling of the form described above is still possible when separate private signals are sent to the Protégé and Target and when the Target may choose to attack the Protégé. It is also shown that even if the Target ignores the private signal from the Third Party, essentially similar equilibria exist in which (1) the Third Party's private signal to the Protégé allows the Protégé to infer the Third Party's type, and (2) the Target learns the Third Party's type from observing the Protégé's behavior. This last result is useful in interpreting cases. The supplementary appendix also analyzes the effects of signaling on the probability of conflict.

## DISCUSSION

The basic signaling mechanism formalized above is simple, and the logic can be easily applied to cases. When a Third Party threatens to defend a Protégé against a Target state, the Third Party will affect the Protégé's behavior toward the Target. Often, the Protégé will adopt a more aggressive or defiant policy vis-à-vis the Target state as a result. These actions may, in turn, increase the likelihood that the Target and Protégé engage in military conflict. As a result, the Third Party's support makes it more likely that the Third Party will actually be faced with the choice of having to follow through on its commitments or risk the destruction of the Protégé. Thus, in Schelling's terms, the Third Party's threat leaves

to chance the possibility that an emboldened Protégé will precipitate a conflict. The Third Party's support shows a willingness to take this chance, causing its statement to convey information to the other states.<sup>7</sup>

If, however, the Protégé has no opportunity to take actions that make conflict more likely or if the Third Party's statements—were they to be believed—would make conflict less likely, then the Third Party's messages would convey nothing to the other players. It is only the increased risk of conflict that sending a statement of commitment entails in equilibrium that causes the statement to convey information about the Third Party's intentions. By incurring a greater risk of war, the Third Party is able to signal to the Target and thereby possibly gain a better negotiated outcome for the Protégé. Thus, entrapment is not just a cost associated with making a strong commitment to an ally. Emboldenment and even the possibility of later entrapment are rather what makes some commitments to allies credible at all.<sup>8</sup>

These considerations imply that this form of signaling is available in some international contexts, but not in all. Two general conditions must hold. First, through emboldenment of the Protégé, support from the Third Party must be expected to increase the likelihood of a disastrous conflict for an unsupported Protégé. Second, the Third Party and Protégé's interests must be sufficiently aligned. This second condition implies that the Third Party would not want to mislead the Protégé into fighting alone; the Third Party must prefer the status quo to a war that the Protégé fights alone.<sup>9</sup>

In the models described above, for the first of these conditions to hold, several subsidiary conditions must hold. The likelihood that the Target is highly resolved must be in a middle range. If the Target is too certain to be resolved, the Protégé would not make a high demand that risks conflict. If the Target is too certain to be of low resolve, the Third Party will be too tempted to support the Protégé even when the Third Party is not resolved to offer aid, and the Protégé

7. Note that in closely related models with a continuum of Third Party cost types, semi-separating equilibria exist in which signals convey information, but bluffing by Third Parties occurs in some cases.

8. Note that in the cheap talk model, messages do not change preferences. Unlike much of the literature relating to entrapment cited above, therefore, the model examines when preferences can be communicated, rather than how preferences can be affected such that a commitment is created. For further analysis of the relationship between emboldenment and entrapment, see the discussion of the related model in supplementary appendix E and, for a partially contrasting perspective, Kim (2011).

9. This is so in the model when  $p < q$  because it is assumed that the Third Party and Protégé have the same preferences over the bargaining space. The condition that  $p < q$  ensures that the Third Party would not want the Protégé to fight when the Third Party would not be willing to come to the Protégé's aid.

will be too tempted to make a high demand whether or not the Third Party offers support. Thus, signaling is impossible when Third Parties have an incentive to overstate their levels of resolve to come to the aid of their Protégés, as in many other models in the literature (Fearon 1995), and also when Third Parties have an incentive to understate their levels of resolve. The Third Party also cannot be too powerful; if it is, the Target will be deterred from risking conflict. When the Third Party sends separate signals to the Protégé and Target, the Protégé must incur an audience cost for backing down from a high demand; otherwise, Third Party support does not embolden the Protégé. Finally, the Protégé must prefer the status quo to fighting alone against the Target; otherwise, the Protégé's actions would not be contingent upon Third Party support.<sup>10</sup>

These considerations can be applied to cases of extended coercion. When Belgium is faced with invasion, for instance, Britain may well prefer that Belgium resist the invasion even without immediate British support because of the potential that an invasion of Britain could be launched from Belgian territory. In this context, therefore, the second general condition for signaling is violated: British and Belgian interests are not so aligned that Britain would be unwilling to risk the destruction of its Protégé when Britain is unwilling or unable to offer immediate support.<sup>11</sup> If British statements of commitment change adversary calculations, therefore, it is not as a result of the signaling logic described here. A key reason for this divergence in interests is the presumption that an invasion would result in the end of Belgium as a political entity. If, on the other hand, a neighboring country were only to threaten to invade a portion of Belgian territory, then Britain might reason that it would be in British interests that Belgium not resist the invasion in order to preserve the rest of Belgian territory—if Britain were in fact not willing to intervene. This dynamic more resembles the negotiations over Czechoslovakia in 1938. If Britain and France had committed to the defense of Czechoslovakia, the Czechs would have fought rather than give up any part of the Sudetenland. Since it was unclear whether the Czech negotiating stance would then have resulted in a conflict with Germany, British and French statements might have sent a strong signal to Germany about British and French resolve. Thus, in this context, both general signaling conditions are satisfied: the British and French believed support would increase the odds

of conflict, and British, French, and Czech interests were sufficiently aligned such that the two Third Parties would not have wanted to commit to support they were not willing to provide.

When a Third Party, along with a Protégé, attempts to compel a Target to take a particular action that the Target may not be willing to take, known as “compellence” (Schelling 1966, 70–71), the Third Party's statements may be quite likely to convey information to the Target. Making such a threat will change the Protégé's behavior toward the Target and, as Schelling (1966) argues, all sides will expect that the potentially humiliating demand that the Target's behavior also change may be resisted. Threats of coordinated action by allies to force behavioral modifications in a third state, therefore, where neither ally would wish to see the other left to face the third state alone and at least one ally is known to be keen to act, will often be credible (Benson 2012). Not all compellent threats will convey information, however. The Target's calculus will not be affected, for instance, when the threat cannot engender a sufficient risk of conflict because the Target is expected to concede, and the Protégé need not take action as a result.

Similarly, some but not all deterrent threats will convey information through the mechanism described here. Suppose, as was the case with negotiations over Czechoslovakia in 1938 for instance, that the Target—Germany—is mobilizing to march to occupy a portion of the territory of the Protégé. We can think of the portion of territory that the Target intends to occupy as represented by  $1 - q$  in the model. The Protégé may only be willing to resist the invasion if the Protégé expects support from the Third Party. In such cases, a statement of support from the Third Party may be able to convey information to the other two states about the Third Party's willingness to fight. In the absence of support from the Third Party, the Target will reject any offer from the Protégé in which the Target does not possess the disputed territory because the Target will understand that any threat of the Protégé's to fight is not credible. On the other hand, a Protégé that expects Third Party support will demand that the invasion cease. If the optimal demand of a Protégé that expects support implies a sizable risk of conflict, then the Third Party will only embolden the Protégé to risk conflict if the Third Party is willing to come to the Protégé's aid if conflict should result. This enables the Third Party to signal its resolve to fight or not to the other states. Since the Third Party's support provides the Protégé with a credible threat to resist an invasion, this support may deter the Target from invading. Note, however, that the Third Party's statements deter only because they *do not* reduce the likelihood of conflict. On the contrary, in order to deter an invasion, the

10. The other specific conditions given in the propositions guarantee that the players are willing to take actions that risk conflict.

11. In the terms of the model, we can think of  $q < p$ , where  $q$  is the portion of territory with which Belgium is left following the invasion. Thus, Britain prefers that Belgium fight without support, but Belgium has the opposite preference.

Third Party must embolden the Protégé in such a way that war may result.

To see how these dynamics operate in the conflict model described above, consider a case where signaling is possible. This occurs in the middle of figure 1, where the probability that the Target is highly resolved,  $h_t$ , is neither too small nor too large, and the probability the Protégé wins a conflict with the Target without support from the Third Party,  $p$ , is not too high. Now, suppose that  $h_t$  increases.<sup>12</sup> If it increases enough, such that the Target is very likely to refuse the Protégé's demand, then the Third Party becomes unwilling to offer support to the Protégé even when the Third Party would be willing to support the Protégé if a war were to result. Equation (3) does not hold; the interests of Third Party and Protégé are not sufficiently aligned. If the Protégé is sufficiently strong (so that  $p$  is sufficiently high) and is therefore willing to precipitate a conflict with the Target, this is a case of *entrapment*. Third Parties that are willing to fight would prefer not to have to, but the Protégé also understands this and will not pay the Third Party's signals any mind. The Protégé will press ahead without the Third Party's commitment of support because the Protégé understands that the lack of support does not mean that the Third Party will not come through when it is faced with the choice of whether or not to allow its Protégé to face the Target alone. The Third Party cannot restrain its own Protégé at the final moment.<sup>13</sup>

Signaling can also fail when  $h_t$  is too low. In fact, signaling must fail when  $h_t$  is sufficiently low because equation (2) cannot hold. The reason is that if it is too likely that the Target will accept a high demand by a Protégé that is offered support, then the Third Party would always choose to offer support, even if it were not willing to really provide it. The Third Party has incentive to overstate its level of resolve because the Target is too likely to back down as a result. Although the interests of Third Party and Protégé are aligned, the Third Party's support does not sufficiently increase the likelihood of conflict. Thus, signaling requires sufficient uncertainty about what the Target will do. Without uncertainty on both sides of a dispute, signals cannot be meaningful. These dynamics are illustrated by the curved line between the Separating and Pooling regions in figure 1.

12. The following parameters produce the equilibrium shown in the figure:  $p^a = .75$ ,  $c_t = c_g = .01$ ,  $\bar{c}_t = .25$ ,  $\underline{c}_d = .2$ ,  $\bar{c}_d = .9$ ,  $h_d = .9$ ,  $q = .7$ . Note that while the figure does not show the whole (0, 1) interval for the parameters  $h_t$  and  $p$ , the illustrated signaling and nonsignaling ranges naturally extend to this interval.

13. Note that for entrapment to be possible, the Protégé must have a sufficiently high expectation,  $h_{d^b}$ , that the Third Party will join in the conflict if a conflict were to begin.

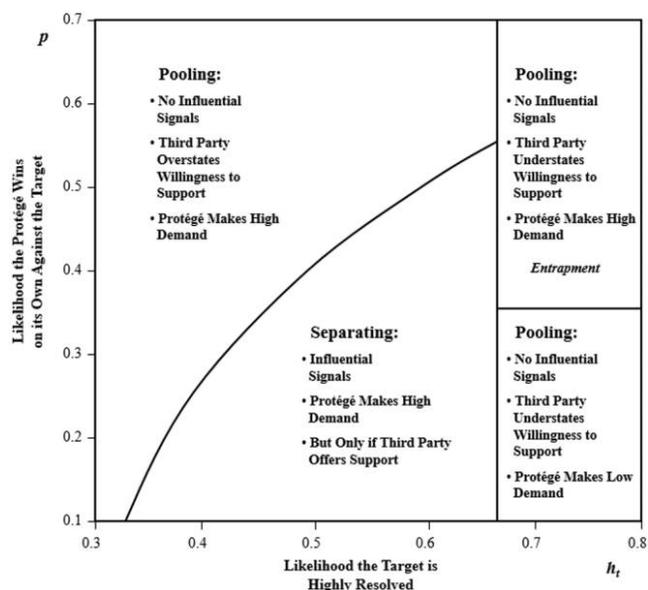


Figure 1. Equilibrium signaling properties

Figure 1 further illustrates that increases in the strength of the Protégé relative to the Target,  $p$ , can also make signaling impossible. The reason is that, when the Protégé is strong, there is less risk to the Third Party in supporting the Protégé when the Third Party is unwilling to come to the Protégé's aid. If the Third Party's support emboldens the Protégé, and the Protégé and Target end up at war as a result, the Third Party reasons, the Protégé still has a good chance of emerging victorious without assistance. This gives the Third Party a powerful incentive to claim to be willing to support the Protégé even when the Third Party isn't. When the Protégé is sufficiently weak, the Third Party has no such incentive. Then, emboldening and not supporting the Protégé incur a real risk of disaster—the destruction of an unsupported Protégé.<sup>14</sup>

As a final observation about the model, note that even though there are many messages that the Third Party could send, there is no equilibrium in which the Third Party can credibly convince the other players that the Third Party will defend the Protégé if and only if the Protégé maintains the status quo and is attacked by the Target. The reason is that

14. Note also that if there is a danger that the Target will attack a pacific Protégé in order to demand more than the status quo, signaling may also be impossible when  $p$  is too low. In the main model described above, for clarity of exposition of the dynamics of the model, a conflict is started if and only if the Protégé elects to begin one. If instead the Target chooses whether or not to begin a conflict, the dynamics are similar as long as the emboldenment of the Protégé outweighs the deterrent effect of the Third Party's statement on the Target so that the overall likelihood of conflict is increased. A model that formalizes this is available in the supplementary appendix.

when a war is win or lose without the possibility of an intermediate outcome, Third Parties either prefer to support the Target in war or prefer to remain out of a conflict. Indeed, in such cases, how could a Third Party acquire a preference to support a Protégé only if the Protégé were attacked? If the Third Party's preferences derive from the objective, material outcome, then no Third Party could acquire such a contingent preference and therefore no such mixture of threats could be credible. Therefore, when a war will result in decisive victory or defeat, if contingent threats are ever to be credible in a rationalist model of politics, the Third Party must have a preference for being listened to, possibly because observers will take note of whether the Third Party was listened to, and this will have implications for the future. Contingent signaling is generally impossible in models of a single crisis, considered in isolation, at least when wars, should they occur, will be decisive.<sup>15</sup>

### THE EMPIRICAL RECORD

As the discussion above makes clear, in general terms, taking for granted that the Protégé and Target are in a bargaining relationship, signaling will be possible when two conditions hold. First, the expectation of the states must be that the increased stridency of the Protégé will increase the likelihood of conflict. Second, the interests of the Third Party and Protégé must be sufficiently aligned. While other conditions are discussed above, such as the requirement that the Third Party's capabilities be in a middle range, these other conditions have the effect of ensuring that these two general conditions hold in the specific model of conflict analyzed.

The central predictions of the model have yet to be tested directly through statistical analysis. While the extensive literatures on extended coercion and the effects of alliances bear most directly on the issues raised here, these studies have not separately examined alliances where the two signaling conditions do and do not hold.<sup>16</sup> Nevertheless, findings in these literatures are broadly consistent with the analysis. Despite a long-running debate about which alliances lead war to be more likely, the literature generally supports the view that some alliances make war more likely, including specifically by encouraging states to resist challenges (Smith 1996). Several studies find an association be-

tween some alliances and the onset of war. Alliances between major powers (Gibler 2000), those that make the balance of power more unequal (Siverson and Tennefoss 1984), those formed for compellence (Benson 2011), in the nineteenth century (Maoz 2000), by dissatisfied states (Gibler 2000; Leeds 2003), by states that were successful in recent wars (Gibler and Vasquez 1998), or in the context of an enduring rivalry between one of the states and a third state (Colaresi and Thompson 2005), have all been found to increase the likelihood of conflict. Also consistent with the findings here, studies have pointed to the importance of a harmony of interests between the Protégé and the Third Party (Danilovic 2002; Huth and Russett 1984).

A qualitative analysis of cases where the conditions for signaling appear to hold can directly investigate whether the implications of the model hold in particular instances, if not the generality of the findings. Therefore, I now turn to the documentary evidence found in a case in which the interests of the Third Party and Protégé were closely aligned, and the states involved therefore understood that the Third Party would not have wished to mislead its Protégé into fighting without support. Thus, the second general signaling condition held. Analysis of additional cases is available in the supplementary appendix.

To evaluate the model, I shall examine whether (1) the Protégé was understood to be emboldened by support (or restrained by its absence), (2) the Third Party's support and the changed behavior of the Protégé were thought to increase the odds of conflict (the first general signaling condition), (3) support or its absence affected the calculations of observers about the Third Party's intentions, and (4) these observers grounded their inferences—at least in part—in whether support would increase the risk of conflict. The method of evaluation will be to analyze the documentary record of the considerations of diplomats and leaders to see if these four predictions are borne out. No argument is made that the mechanism described here exerted exclusive influence. Indeed, it is clear that decision makers also considered other factors (Press 2005). Rather, the historical data will be considered to constitute evidence for the signaling mechanism if the documentary record shows diplomats and leaders considering the situations in a manner consistent with these four aspects of the model.

### British support for Czechoslovakia and Poland in 1938–39

In the record of British diplomacy prior to the Second World War, we find direct evidence for the considerations and inferences predicted by the model. Negotiations between Czechoslovakia and Germany initially focused on the

15. Contingent and probabilistic alliance commitments are examined, both theoretically and empirically, in Benson (2011, 2012).

16. Literature on extended deterrence analyzes when deterrent threats are likely to be coercive. Fearon (1994b) and Huth (1988a), for instance, find mixed evidence that threats make Targets more likely to back down, but coercive success can be quite different from convincing an adversary.

treatment of Sudeten Germans. Negotiations between Poland and Germany centered on the treatment of Germans within Poland and whether Germany would be allowed to annex Danzig and the “Corridor” between that city and the German state. Germany had no intention of allowing the negotiations with either Czechoslovakia or Poland to reach a conclusion that did not involve annexation and was therefore intensely interested in understanding what the British would do in the event of German action.

Britain supported Czechoslovakia and Poland, but British support was not as unequivocal as it might have been. In the case of Czechoslovakia, Britain eventually signaled that it would permit Germany to annex the Sudetenland, and the Czechs acquiesced, believing Britain’s signal that it would not join in a conflict to be credible. In the case of Poland, Britain expressed a clear commitment to defend Polish independence, but not necessarily Polish integrity, a distinction that was noted by the Germans. German leaders drew conclusions both from British support and from Britain’s lack of full commitment.

All sides understood that British support would embolden its Protégés. According to one German memorandum, “The Czechs will only become reasonable if the British and, by their agency, the French, express their intention of not sacrificing soldiers’ lives for the stupidities of Czech policy. The Czech press, including the Beneš press, was persuading the Czechs that Britain would help in any contingency. It must plainly be stated that she would not help.”<sup>17</sup> The German understanding of the situation in negotiations over Poland was similar. Hitler argued that Poland would not have resisted the transfer of Danzig without British support (Watt 1989, 317).

For their part, the British saw things in much the same way. They understood that supporting their protégés would embolden them and thereby increase the likelihood of full-scale conflict. Support for Poland, for example, would be a signal of British resolve to defend Poland in part because British support was expected by all sides to embolden Poland in negotiations over Danzig and the Corridor and, therefore, to make military conflict more likely. The evidence from British documents is explicit: “We are not in a position to assess the deterrent effect of such a Pact upon Germany, but an important military implication is that if such a Pact were to encourage an intransigent attitude on the part of Poland and Romania, it would thereby tend to precipitate a European war before our forces are in any way fully prepared for it, and such a war might be started by

17. Documents on German Foreign Policy, Series D, vol. 2, p. 546, memorandum, August 10, 1938.

aggression against Danzig alone” (Alexandroff and Rosecrance 1977, 411). A stronger documentary confirmation of the expectation of support leading to the increased likelihood of conflict could hardly be expected.

In fact, precisely because he worried that emboldening Poland might lead to conflict, Halifax tried to restrain the Polish government. He wrote to them in the heat of the crisis: “I should not expect the Polish Government to abandon all hope of negotiation unless they were convinced that it afforded no possibility of averting a threat to Polish independence . . . if [they] wished to establish that there ‘clearly’ was such a threat, they would naturally desire to consult with His Majesty’s Government and would therefore do so *before* taking any irrevocable action.” (Thorne 1968, 159–60). There is even substantial evidence that Britain was in fact willing to give up Danzig and the Corridor and perhaps even all of Poland under some circumstances (Shore 2002, 87–100). Thus, the British did not offer unequivocal private support to Poland. According to the analysis above, this lack of support for the integrity of Poland should have influenced Polish behavior and German inferences.

Indeed, as expected, Germany saw the degree of intransigence of the Poles in negotiations as a direct indicator of the strength of the British commitment to Poland. On the one hand, Hitler believed Poland would not have protested the transfer of Danzig without a measure of British support. On the other hand, the Germans (along with the British) realized that Poland would have been even less willing to negotiate if British support had been stronger (Watt 1989, 185). In discussions in mid-August, Hitler argued that Britain would not defend Poland because if Britain were truly supporting Poland, the latter would have been more “cocky.”<sup>18</sup> Had Britain been less equivocal in its support of Poland, it is likely that Germany would have been convinced of British resolve sooner. Thus, at each step, the considerations of the senior German officials support the analysis above. All sides believed that diplomatic support—when it was offered—emboldened, and all sides believed emboldenment increased the likelihood of conflict. Further, the documents show that German elites specifically grounded their inference that Britain would not intervene in Poland in the argument that Poland would have been further emboldened, leading to an increased risk of conflict, if Poland had received full British support.

18. Documents on German Foreign Policy, Series D, vol. VII, p. 555. Hitler also specifically noted that Britain had declined to offer Poland financial assistance and reasoned that, “This suggests that England does not really want to support Poland.” Documents on German Foreign Policy, Series D, vol. VII, pp. 203, 553.

**CONCLUSION**

The benefits of signaling resolve to third parties come at a cost, namely, increasing the probability of war by increasing the assertiveness of allies. Just because signals are credible doesn't make them worth sending. In fact, as in other models in the literature, the higher the risk and more severe the drawbacks, the more likely the signal will be found credible.<sup>19</sup> The implication for policy making is not that particular courses of action be taken, but for policy makers to be aware of the trade-offs. As in other areas, the policies that appear the most successful with the benefit of hindsight will often be the policies that entailed the greatest risks at the time.

An interesting implication of the analysis is that, if the Third Party's signals are to convey information, neither the Third Party nor the Protégé can be too strong. Too powerful Third Parties have too much incentive to make threats on which they would not be willing to follow through because even relatively resolved Targets would make dramatic concessions to avoid conflict. Similarly, if the Protégé's prospects in a war it fights alone are too favorable, Third Parties have no disincentive to overstate their resolve—the Protégé will be relatively well off on its own. However, if the Third Party and Protégé are not sufficiently formidable when fighting in combination, the Third Party's statements will again have no effect on the other players' beliefs.

The mechanism of diplomatic inference described here sheds light on signaling in a range of substantive areas that are often examined independently. The model illuminates new dynamics in extended deterrence when the interests of Third Party and Protégé are aligned and the Protégé and Target are negotiating over issues that could result in conflict. The same prediction applies to compellence cases, where the Third Party and Protégé may seek to coerce the Target together. This means of drawing inferences is never the sole avenue by which diplomats and leaders draw conclusions about each other's intentions, but the documentary evidence across a range of cases for the dynamics described in the model is substantial. It is in part through these processes that states form the expectations about who will side with whom that inform calculations of national interest. The effects of such signals are thus likely to be felt past the resolution of a particular diplomatic episode or crisis and to influence the construction of the international order of the day.

19. For instance, in Fearon (1994a), audience costs imply that the weaker a state and the more likely its adversary is to stand firm, the more information a threat that precipitates a crisis conveys about the state's resolve.

**APPENDIX**

**PROOF OF PROPOSITION 1**

I will show that under the conditions given in the proposition, the following strategies and beliefs constitute a perfect Bayesian equilibrium. The Third Party's strategy is: if  $c_d = \underline{c}_d$ , send  $m = 1$  and fight; if  $c_d = \bar{c}_d$ , send  $m = 0$  and don't fight. The Target's strategy is: following  $m = 1$ , accept the Protégé's offer  $x$  iff  $x \leq p^a + c_i$ ; following any other message  $m$ , accept the Protégé's offer  $x$  iff  $x \leq q$ . The Protégé's strategy is: following  $m = 1$ , demand  $x = p^a + \bar{c}_i$ ; following any other message  $m$ , demand  $q$ ; fight following accepted offers iff  $x < \mu_g p^a + (1 - \mu_g)p - c_g$  and following rejected offers iff  $q < \mu_g p^a + (1 - \mu_g)p - c_g$ . Let  $\mu_t(m, x)$  represent the Target's updated beliefs following  $m$  and  $x$  that  $c_d = \underline{c}_d$ ;  $\mu_t$  will refer to this function in the proofs. The Protégé and Target's updated beliefs are  $\mu_g(1) = \mu_t(1, p^a + \bar{c}_i) = 1$ ,  $\mu_g(m) = 0 \forall m \neq 1$ ,  $\mu_t(m, x) = 0 \forall m \neq 1, x \neq p^a + \bar{c}_i$ . We need not specify the players' updated beliefs about the Target's type following the Target's decision because player choices do not depend on these beliefs.

The optimality of the Target's strategy follows directly from its preferences and equilibrium beliefs, and the players' beliefs follow directly from Bayes's rule and the Third Party's strategy. To see that the Protégé's strategy is optimal, note that given its beliefs and the Target's strategy, following  $m = 1$ , the Protégé must demand either  $p^a + \underline{c}_i$  or  $p^a + \bar{c}_i$ . Since  $p^a + \underline{c}_i$  is accepted with certainty, a lesser demand yields a worse outcome for the Protégé. Any demand  $x \in (p^a + \underline{c}_i, p^a + \bar{c}_i)$  incurs the same risk of conflict as  $p^a + \bar{c}_i$  but yields a worse outcome if accepted. Any demand above  $p^a + \bar{c}_i$  is sure to be rejected, which leads to a worse outcome than  $p^a + \underline{c}_i$ . Following the logic given in the text, for  $p^a + \bar{c}_i < 1$ , the Protégé prefers the higher demand when equation (1) holds. Call the RHS (right-hand side) of equation (1)  $\hat{h}_i$ . Announcing  $q$  following  $m \neq 1$  is optimal because any other strategy yields at most  $q$  given the players' strategies and beliefs. The optimality of the Protégé's war choice follows directly from its preferences and updated beliefs. The optimality of the Third Party's choice at its final choice node follows directly from its preferences. Given the player's strategies and beliefs, which generate the expected utilities given in the text, the Third Party of type  $c_d = \underline{c}_d$  prefers to send a message if equation (3) holds. Call the RHS of equation (3)  $\tilde{h}_i$ . Let  $\bar{h}_i = \min\{\hat{h}_i, \tilde{h}_i\}$ . Similarly, the Third Party of type  $c_d = \bar{c}_d$  prefers to send message  $m = 0$  if equation (2) holds. Call the RHS of equation (2)  $\underline{h}_i$ .

For sufficiently low  $p < q$ ,  $\underline{h}_i < 1$ . Thus, for sufficiently low  $\underline{c}_i$  and  $c_g$ ,  $\hat{h}_i > \underline{h}_i$ . Similarly, for  $\underline{c}_d$  sufficiently low,  $\tilde{h}_i > \underline{h}_i$ . Thus, for  $h_i \in (\underline{h}_i, \bar{h}_i)$ , the Protégé's strategy is optimal, as are the message choices of both Third Party types.

## ACKNOWLEDGMENTS

I am grateful to Robert Jervis, Andy Kydd, Chad Nelson, Marc Trachtenberg, the *JOP* editors, and to several extraordinarily helpful external reviewers, particularly to one whose comments led to dramatic revisions of the article.

## REFERENCES

- Alexandroff, Alan, and Richard Rosecrance. 1977. "Deterrence in 1939." *World Politics* 29 (3): 404–24.
- Benson, Brett V. 2011. "Unpacking Alliances: Deterrent and Compellent Alliances and Their Relationship with Conflict, 1816–2000." *Journal of Politics* 73:1111–27.
- Benson, Brett V. 2012. *Constructing International Security: Alliances, Deterrence, and Moral Hazard*. Cambridge: Cambridge University Press.
- Benson, Brett V., and Emerson M. S. Niu. 2005. "Public Opinion, Foreign Policy, and the Security Balance in the Taiwan Strait." *Security Studies* 14 (2): 274–89.
- Cetinyan, Rupen. 2002. "Ethnic Bargaining in the Shadow of Third-Party Intervention." *International Organization* 56 (3): 645–77.
- Clare, Joe, and Vesna Danilovic. 2010. "Multiple Audiences and Reputation Building in International Conflicts." *Journal of Conflict Resolution* 54 (6): 860–82.
- Colaresi, Michael P., and William R. Thompson. 2005. "Alliances, Arms Buildups and Recurrent Conflict: Testing a Steps-to-War Model." *Journal of Politics* 67 (2): 345–64.
- Crawford, Timothy J. 2003. *Pivotal Deterrence: Third-Party Statecraft and the Pursuit of Peace*. Ithaca, NY: Cornell University Press.
- Crawford, Vincent P., and Joel Sobel. 1982. "Strategic Information Transmission." *Econometrica* 50 (6): 1431–51.
- Danilovic, Vesna. 2002. *When the Stakes Are High*. Ann Arbor: University of Michigan Press.
- Fang, Songying, Jesse C. Johnson, and Brett Ashley Leeds. 2014. "To Concede or to Resist? The Restraining Effect of Military Alliances." *International Organization* 68:775–809.
- Fearon, James D. 1994a. "Domestic Political Audiences and the Escalation of International Disputes." *American Political Science Review* 88 (3): 577–92.
- Fearon, James D. 1994b. "Signaling versus the Balance of Power and Interests: An Empirical Test of a Crisis Bargaining Model." *Journal of Conflict Resolution* 38 (2): 236–69.
- Fearon, James D. 1995. "Rationalist Explanations for War." *International Organization* 49 (3): 379–414.
- Forges, Françoise. 1990. "Universal Mechanisms." *Econometrica* 58 (6): 1341–64.
- Gartner, Scott Sigmund, and Randolph M. Siverson. 1996. "War Expansion and War Outcome." *Journal of Conflict Resolution* 40 (1): 4–15.
- Gaubatz, Kurt Taylor. 1996. "Democratic States and Commitment in International Relations." *International Organization* 50 (1): 109–39.
- Gibler, Douglas M. 2000. "Alliances: Why Some Cause War and Why Others Cause Peace." In John A. Vasquez, ed., *What Do We Know about War*. Lanham, MD: Rowman & Littlefield.
- Gibler, Douglas M., and John A. Vasquez. 1998. "Uncovering the Dangerous Alliances, 1495–1980." *International Studies Quarterly* 42 (4): 785–807.
- Goldstein, Avery. 2000. *Deterrence and Security in the 21st Century: China, Britain, France, and the Enduring Legacy of the Nuclear Revolution*. Stanford, CA: Stanford University Press.
- Grigoryan, Arman. 2010. "Third-Party Intervention and the Escalation of State-Minority Conflicts." *International Studies Quarterly* 54 (4): 1143–74.
- Guisinger, Alexandra, and Alastair Smith. 2002. "Honest Threats: The Interaction of Reputation and Political Institutions in International Crises." *Journal of Conflict Resolution* 46 (2): 175–200.
- Huth, Paul. 1988a. "Extended Deterrence and the Outbreak of War." *American Political Science Review* 82 (2): 423–43.
- Huth, Paul. 1988b. *Extended Deterrence and the Prevention of War*. New Haven, CT: Yale University Press.
- Huth, Paul, and Bruce Russett. 1984. "Testing Deterrence Theory: Rigor Makes a Difference." *World Politics* 42 (4): 270–90.
- Jervis, Robert. 1970. *The Logic of Images in International Relations*. New York: Columbia University Press.
- Jervis, Robert, Richard Ned Lebow, and Janice Gross Stein. 1985. *Psychology and Deterrence*. Baltimore, MD: Johns Hopkins University Press.
- Karsten, Peter, Peter D. Howell, and Artis Frances Allen. 1984. *Military Threats: A Systematic Historical Analysis of the Determinants of Success*. Westport, CT: Greenwood.
- Kilgour, Mark D., and Frank C. Zagare. 1994. "Uncertainty and the Role of the Pawn in Extended Deterrence." *Synthese* 100 (3): 379–412.
- Kim, Tongfi. 2011. "Why Alliances Entangle but Seldom Entrap States." *Security Studies* 20 (3): 350–77.
- Kuperman, Alan J. 2008. "The Moral Hazard of Humanitarian Intervention: Lessons from the Balkans." *International Studies Quarterly* 52 (1): 49–80.
- Kurizaki, Shuhei. 2007. "Efficient Secrecy: Public versus Private Threats in Crisis Diplomacy." *American Political Science Review* 101 (3): 543–58.
- Kydd, Andrew. 1997. "Game Theory and the Spiral Model." *World Politics* 49 (3): 371–400.
- Lebow, Richard Ned, and Janice Gross Stein. 1989. "Rational Deterrence: I Think, Therefore I Deter." *World Politics* 41 (2): 208–24.
- Leeds, Brett Ashley. 1999. "Domestic Political Institutions, Credible Commitments, and International Cooperation." *American Journal of Political Science* 43 (4): 979–1002.
- Leeds, Brett Ashley. 2003. "Do Alliances Deter Aggression? The Influence of Military Alliances on the Initiation of Militarized Interstate Disputes." *American Journal of Political Science* 47 (3): 427–39.
- Maoz, Zeev. 1983. "Resolve, Capabilities, and the Outcomes of Interstate Disputes, 1816–1976." *Journal of Conflict Resolution* 27 (2): 195–229.
- Maoz, Zeev. 2000. "Alliances: The Street Gangs of World Politics—Their Origins, Management, and Consequences, 1816–1986." In John A. Vasquez, ed., *What Do We Know about War*. Lanham, MD: Rowman & Littlefield.
- Morgan, Patrick M. 1983. *Deterrence: A Conceptual Analysis*. Cambridge: Cambridge University Press.
- Morrow, James D. 1994. "Alliances, Credibility, and Peacetime Costs." *Journal of Conflict Resolution* 38:270–97.
- Morrow, James D. 2000. "Alliances: Why Write Them Down?" *Annual Review of Political Science* 3 (1): 63–83.
- O'Hanlon, Michael. 2000. "Why China Cannot Conquer Taiwan." *International Security* 25 (2): 51–86.
- Powell, Robert. 1996a. "Bargaining in the Shadow of Power." *Games and Economic Behavior* 15:255–89.
- Powell, Robert. 1996b. "Stability and the Distribution of Power." *World Politics* 48 (2): 239–67.
- Press, Daryl G. 2005. *Calculating Credibility: How Leaders Assess Military Threats*. Ithaca, NY: Cornell University Press.
- Quackenbush, Stephen L. 2006. "Not Only Whether but Whom: Three-Party Extended Deterrence." *Journal of Conflict Resolution* 50 (4): 562–83.
- Russett, Bruce M. 1963. "The Calculus of Deterrence." *Journal of Conflict Resolution* 7:97–109.

- Sartori, Anne E. 2005. *Deterrence by Diplomacy*. Princeton, NJ: Princeton University Press.
- Schelling, Thomas C. 1966. *Arms and Influence*. New Haven, CT: Yale University Press.
- Schultz, Kenneth. 2001. *Democracy and Coercive Diplomacy*. Cambridge: Cambridge University Press.
- Sechser, Todd S. 2010. "Goliath's Curse: Coercive Threats and Asymmetric Power." *International Organization* 64 (4): 627–60.
- Shore, Zachary. 2002. *What Hitler Knew: The Battle for Information in Nazi Foreign Policy*. New York: Oxford University Press.
- Signorino, Curtis S., and Ahmer Tarar. 2006. "A Unified Theory and Test of Extended Immediate Deterrence." *American Journal of Political Science* 50 (3): 586–605.
- Siverson, Randolph M., and Michael R. Tennefoss. 1984. "Power, Alliance, and the Escalation of International Conflict, 1815–1965." *American Political Science Review* 78 (4): 1057–69.
- Smith, Alastair. 1995. "Alliance Formation and War." *International Studies Quarterly* 39 (4): 405–25.
- Smith, Alastair. 1996. "To Intervene or Not to Intervene." *Journal of Conflict Resolution* 40 (1): 16–40.
- Smith, Alastair. 1998a. "Extended Deterrence and Alliance Politics." *International Interactions* 24 (4): 315–43.
- Smith, Alastair. 1998b. "International Crises and Domestic Politics." *American Political Science Review* 92 (3): 623–38.
- Snyder, Glenn H. 1997. *Alliance Politics*. Ithaca, NY: Cornell University Press.
- Snyder, Glenn Herald, and Paul Diesing. 1977. *Conflict among Nations: Bargaining, Decision Making, and System Structure in International Crises*. Princeton, NJ: Princeton University Press.
- Thorne, Christopher. 1968. *The Approach of War, 1938–39*. London: Macmillan.
- Tomz, Michael. 2007. "Domestic Audience Costs in International Relations: An Experimental Approach." *International Organization* 61 (4): 821–40.
- Trager, Robert F. 2010. "Diplomatic Calculus in Anarchy: How Diplomacy Matters." *American Political Science Review* 104 (2): 347–68.
- Trager, Robert F. 2011. "Multi-Dimensional Diplomacy." *International Organization* 65:469–506.
- Trager, Robert F. 2013. "How the Scope of a Demand Conveys Resolve." *International Theory* 5 (3): 414–45.
- Trager, Robert F., and Lynn Vavreck. 2011. "The Political Costs of Crisis Bargaining: Presidential Rhetoric and the Role of Party." *American Journal Political Science* 55 (3): 526–45.
- Wagner, R. Harrison. 2004. "Bargaining, War, and Alliances." *Conflict Management and Peace Science* 21 (3): 215–31.
- Walter, Barbara F. 2002. *Committing to Peace: The Successful Settlement of Civil Wars*. Princeton, NJ: Princeton University Press.
- Watt, Donald Cameron. 1989. *How War Came: The Immediate Origins of the Second World War, 1938–1939*. London: Heinemann.
- Werner, Suzanne. 2000. "Deterring Intervention: The Stakes of War and Third-Party Involvement." *American Journal of Political Science* 44 (4): 720–32.
- Yuen, Amy. 2009. "Target Concessions in the Shadow of Intervention." *Journal of Conflict Resolution* 53 (5): 745–73.
- Zagare, Frank C., and Mark D. Kilgour. 2003. "Alignment Patterns, Crisis Bargaining, and Extended Deterrence." *International Studies Quarterly* 47 (4): 587–615.
- Zartman, William, and Guy Faure. 2005. *Escalation and Negotiation in International Conflicts*. New York: Cambridge University Press.