The Informational Role of International Institutions and Domestic Politics

Songying Fang University of Minnesota

Why did President Bush attempt to acquire a UN Security Council resolution authorizing the use of force before the war with Iraq, even though there was a substantial risk that his request would be rejected? This article presents a game-theoretic model to investigate how international institutions can shape the behavior of democratic leaders by influencing domestic politics. While it seems unsurprising that unbiased leaders who are truly concerned about foreign policy outcomes would consult international institutions, the results show that biased leaders with private agendas can also be forced to behave like the unbiased type because of their electoral concerns. The equilibrium results are illustrated with the cases of U.S. use of force in international crises.

In a public opinion poll conducted between January 21 and 26, 2003, 67% of Americans surveyed responded that it was necessary for the United States to get approval from the UN Security Council to invade Iraq. Furthermore, three polls taken between June 2002 and January 2003 showed that there was a continuing insistence on UN approval of U.S. military action by a majority of the Americans (PIPA/Knowledge Networks 2003a). These poll results are puzzling considering that some of the veto players of the UN Security Council are not particularly well perceived by the American public. Few Americans surveyed would respond that acquiring consent from Russia and China would be necessary to justify a war against Iraq; yet, when the question was phrased as acquiring an approval from the UN Security Council, a solid majority answered that a joint endorsement from these countries is necessary. As if responding to the public sentiment, the Bush administration went to considerable length to acquire an authorization from the Security Council before initiating the war.[1] The episode suggests that the endorsement of an international institution means something to domestic audiences and that state leaders could be compelled to seek support from such institutions because of domestic concerns.

The linkage between international institutions and state behavior has long been recognized (Keohane 1984). In recent years, as the literature on international institutions sought to move beyond the initial question of *whether* such institutions matter, scholars have begun to examine the causal mechanisms underlying the relationship (Martin and Simmon 1998). Importantly, studies in the area of international political economy have identified domestic politics as the driving force behind the executives' incentives to join and comply with international agreements (Dai 2005; Mansfield, Milner, and Rosendorff 2002). The thrust of the argument is that voters and domestic interest groups keep a watchful eye on their leaders' involvement in trade and environmental agreements and hold them accountable for their policy choices; consequently, leaders display more cooperative behavior than they would otherwise. What is not obvious, however, is whether such a mechanism exists for international conflict, where leaders and domestic groups feel most strongly about their national autonomy and international institutions are expected to have the least impact (Grieco 1988; Mearsheimer 1994–95). In addition, in a highly uncertain environment of an international crisis, voters may not have a good sense of which policy serves their interests

Songying Fang is assistant professor of political science, University of Minnesota, 1414 Social Sciences Building, Minneapolis, MN 55455 (sfang@umn.edu).

I wish to thank John Duggan, Curtis Signorino, Randall Stone, Dan Bernhardt, Mark Fey, John Freeman, Yue Fu, Hein Goemans, Guishu Jin, Jim Johnson, Mark Souva, and Erik Voeten for valuable comments and suggestions. Special thanks to Seok-ju Cho, Justin Fox, and Kris Ramsay for many discussions. I would also like to thank the editor of the *AJPS* and three anonymous reviewers. All errors are the sole responsibility of the author.

[1] I discuss the circumstances under which President Bush appealed to the UN Security Council in more detail later on in the context of the equilibrium results.

American Journal of Political Science, Vol. 52, No. 2, April 2008, Pp. 304–321

 ISSN 0092-5853

best, and thus face a greater challenge in evaluating their leader's policy than if they were dealing with trade and environmental issues. How may international institutions help resolve this sort of inferential problems for voters? I investigate these questions with a game-theoretic model that explicitly incorporates domestic politics into the interaction between states and international institutions.

The model is developed around the idea that international institutions are a source of information for various audiences,[2] highlighting in particular, the signaling effect of the institutions in the presence of asymmetric information between leaders and voters (Mansfield, Milner, and Rosendorff 2002; Simmons 2000). There are two types of leaders in the model: biased or unbiased. Both types draw utility from policy outcome and staying in office; however, a biased leader also receives a private benefit if she chooses a particular policy. Confronted with an international crisis and incomplete information about the state of the world, the leader can either take a unilateral action or consult an international institution before choosing a policy. In the election that follows the crisis, the median voter decides whether or not to keep the incumbent in office based on his updated belief about the leader's type.

The equilibrium results suggest that there is no fully separating equilibrium in which one type of leader always consults the institution and the other type only takes unilateral actions. In particular, if unbiased leaders consult the institution in equilibrium, then biased leaders will be forced to behave like the unbiased type to some extent by taking an issue to the institution because of their electoral concerns. After initially consulting the institution, however, biased leaders may behave differently from the unbiased type depending on the private benefit they receive from a particular policy. If the benefit is small, then there is a unique pooling institutional equilibrium in which a biased leader fully mimics the behavior of the unbiased type and follows what the institution recommends. If the benefit is larger, then *partial* pooling behavior occurs in equilibrium: the biased type mimics the unbiased type in consulting the institution, but reveals her type subsequently by taking her preferred action if the institution advises otherwise.

The analysis contributes to the studies of international institutions as well as international conflict. First, the model provides a microfoundation for the conflict-reducing effect of international institutions. A series of empirical studies have found that such an effect exists (Mitchell and Hensel 2007; Oneal and Russett 2003; Russett, Oneal, and Davis 1998); however, there is a lack of theoretical work that clearly identifies the underlying mechanisms. Without assuming that an international institution has enforcement power, I show that such an institution can nevertheless have the effect of disciplining leaders with private agendas through the electoral process. Second, the article continues in the tradition of opening up the "black box" of domestic politics, but focusing on a relatively unexplored area—the interaction between domestic politics and international institutions. Although there is an extensive literature that looks into the effects of democratic politics on international conflict and trade, research incorporating the role of international institutions in the relationship has only begun to emerge (Chapman 2007; Dai 2005; Mansfield, Milner, and Rosendorff 2002). In the next section I elaborate on a few features of the model that are related to but also substantially different from the existing studies.

Elections, Information, and International Institutions

Several aspects of democratic politics have been recognized as influencing leaders' behavior in international events, including legislative veto power (Milner and Rosendorff 1997; Mo 1995; Putnam 1988; Tarar 2001), political opposition (Ramsay 2004; Schultz 1998), and the electoral process (Dai 2005; Fearon 1994; Guisinger and Smith 2002; Mansfield, Milner, and Rosendorff 2002; Smith 1998). As the literature aiming at understanding the effect of domestic politics develops, how incomplete information interacts with leaders' incentives to behave strategically in front of different audiences has become a central concern. The issue of asymmetric information is particularly salient in the context of international crises and subsequent elections, since ordinary voters have neither the access to the critical information possessed by a leader (Canes-Wrone, Herron, and Shotts 2001; Coate and Morris 1995; Mansfield, Milner, and Rosendorff 2002), nor the opportunity to evaluate the quality of a leader through close contact. How do we capture the inferential problems voters face in such an environment?

Because voters have a fundamental interest in electing a leader who brings the best policy outcome, a straightforward way of modeling the voting decision is to assume that voters use policy outcome as an indicator of a leader's quality that is not directly observable (Smith 1998). The approach captures an important aspect of a leader's

[2]Numerous authors have made the observation. For recent works, see Haftendorn, Keohane, and Wallander (1999), Mitchell (1998), and Dai (2002).

quality, namely a leader's ability to execute a policy successfully. However, it leaves out a prior question of whether the policy chosen is appropriate—a good policy outcome results from both the choice of a sound policy and its successful execution.[3] Moreover, international conflicts are typically resolved over a long period of time, and a policy outcome may not be fully observed at the time of a reelection (Holsti 2004, 195). A natural extension of the existing approach, therefore, is to assume that voters wish to reelect *unbiased* leaders who care about choosing appropriate policies. This is a critical assumption employed in the model, and it leads to important new insights about the signaling effect of international institutions.

Overwhelming evidence suggests that voters are on balance not well informed about foreign affairs (Holsti 2004, 51). In line with the finding, the model does not require that the voter *knows* what the correct policy is to recognize an unbiased leader; rather, it assumes that the voter can make use of the fact that different types of leaders have different incentives in the presence of an international institution. In particular, an unbiased leader is more inclined to seek information from the institution to assist her policy choice, while a biased leader has an incentive to avoid consulting such an institution when she is predisposed to a policy. The asymmetric information between the leader and the voter provides an opportunity for the institution to influence the voter's decision by providing a signaling device for the unbiased type.[4]

What kind of information can an international institution provide for domestic audiences? An international institution may have informational advantage over states on technical issues, but it can also provide information on "the interest, preference and intentions of other states" (Haftendorn, Keohane, and Wallander 1999, 4) by forcing states to announce their positions publicly through deliberation and voting. Such information can be used by the audiences to assess the level of international support (or opposition) for a policy and the likelihood that the policy will succeed.[5] The information may not reach voters, however. It is conceivable that while some leaders turn to international institutions for more information, others may avoid such a course of action since the information revealed by an institution could contradict what they would like the public to believe. One of the important features of the model is that it allows the leader to have a choice of soliciting or blocking the information from the institution. This is an important departure from the existing studies, which often implicitly assume that states always prefer more information. Allowing leaders to have the choice puts a significant constraint on the informational role of international institutions; nevertheless, the model predicts that the existence of such an institution nontrivially changes leaders' behavior through domestic politics.

The Model

Suppose there is a democratic state involved in an international crisis. Let P denote the leader and v the median voter. Facing the crisis, P decides what foreign policy to adopt, and the voter decides if P will remain in office after P's policy is implemented. There is also an international institution, I, that gives recommendations on foreign policy issues if it is consulted by the leader.

The leader cares about policy outcome as well as getting reelected. Let $x \in \{0, 1\}$ denote P's policy. P could choose and implement a policy unilaterally, or she could consult I before choosing a policy. There are two possible states of the world, $\omega \in \{0, 1\}$. Assume that it is equally likely that $\omega = 0$ or $\omega = 1$.[6] A policy that matches the state of the world ($x = \omega$) is called a *correct* (or appropriate) policy. A correct policy succeeds with probability $\alpha \in (0, 1)$, while an incorrect policy fails with probability 1.[7] Let $z \in \{0, 1\}$ denote the outcome of a policy. If the policy is a success, then $z = 1$; if the policy is a failure, then $z = 0$.

The leader has two types that differ in their policy preferences. If P is unbiased, then P's only concern with respect to policy is to choose a correct one; if P is biased, then choosing policy $x = 1$ brings her a private benefit.[8] Let $\gamma \in \{0, b\}$ denote P's type: $\gamma = 0$ if P is unbiased, and $\gamma = b$ if P is biased, where $b > 0$ measures the private

[3] A casual observation of the debate surrounding the current war in Iraq suggests that voters do care about the appropriateness of a policy as much as its execution (Kull 2004, 2006).

[4] In their empirical study of U.S. use of force, Chapman and Reiter (2004) find that the rally 'round the flag effect is bigger when the UN Security Council supports the use of force than otherwise.

[5] For example, Voeten (2005) suggests that UN Security Council decisions signal whether a use of force violates an international norm that should be defended.

[6] This assumption can be relaxed to one that assumes one state of the world is more likely than the other. It can be shown formally that while the specific condition in each proposition will change accordingly, the nature of the result remains the same.

[7] The assumption that an incorrect policy fails with probability 1 is made to simplify the algebra. All the results hold with a relaxed assumption that an incorrect policy fails with some high probability.

[8] It is not uncommon in political economy literature to assume that politicians have different types and some are more susceptible to policy biases (e.g., Coate and Morris 1995).

benefit from policy $x = 1$ for a biased leader. Both types of leaders want to get reelected. Let $r \in \{0, 1\}$ denote how the voter votes in a reelection following an international crisis: if $r = 1$, then P gets reelected; if $r = 0$, then P is replaced. To summarize, the leader's utility function, $u_P(\gamma, x, z, r)$, depends on the leader's type, the policy chosen, the outcome of the policy, and the reelection result:

$$u_P(\gamma, x, z, r) = \gamma x + z + r.$$

We can see that while both types of leaders draw utility from the *outcome* of a policy, only the biased type ($\gamma = b$) draws utility from policy $x = 1$ per se.

The institution's preference is to maximize the probability of policy success. This is clearly a simplifying assumption to focus our attention on the interaction between the leader and the voter; however, it is worth noting that if an institution is known to be always biased toward a certain type of policy, then it will lose its credibility as a source of information with an audience. It is precisely when an institution allows its decision to be contingent on the information that it receives that leaders will pay any attention to what it says. Therefore, by assuming that the institution cares about policy success, the model examines the cases where other political considerations do not overwhelm an institution's interest in seeing an appropriate policy being executed. The utility function of the institution, $u_I(z)$, thus depends simply on the outcome of P's policy:

$$u_I(z) = z.$$

Let $d \in \{0, 1\}$ denote I's recommendation to P, where $d = 0$ corresponds to recommending policy $x = 0$, and $d = 1$ corresponds to recommending policy $x = 1$.

In the election that follows the crisis, the voter's concern is to reelect an unbiased leader and replace a biased leader. The utility function of the voter thus depends on whether or not he voted correctly. Assume that the voter's belief at the beginning of the game is that P is type $\gamma = 0$ with probability $\pi_0 > 1/2$, and type $\gamma = b$ with probability $1 - \pi_0$. The voter will update his belief after watching the interaction between the leader and the institution. The key to the update is the voter's knowledge of the equilibrium behavior of different types of leaders. For example, if in equilibrium unbiased leaders are more likely to consult the institution and then follow the institution's advice, then the voter will update favorably about a leader's type when seeing such behavior. As another example, if in equilibrium unbiased leaders are more likely to take unilateral action $x = 0$, while biased leaders are more likely to take unilateral action $x = 1$, then the voter will only update favorably about a leader's type seeing action $x = 0$.

Let $u_v(r, \gamma)$ denote the voter's utility based on his voting decision. If the voter reelected an unbiased leader, then $u_v(1, 0) = 1$; if the voter reelected a biased leader, then $u_v(1, b) = 0$.[9] On the other hand, if the voter voted to replace an incumbent, a lottery results: the new leader is an unbiased type with probability π_0 and a biased type with probability $1 - \pi_0$, where π_0 is the prior belief about a leader's type. Thus, the voter's utility function can be summarized as follows:

$$u_v(r, \gamma) = \begin{cases} u_v(1, 0) = 1 \\ u_v(1, b) = 0 \\ u_v(0, 0) = u_v(0, b) = \pi_0 \end{cases}$$

As discussed earlier, the voter does not judge the quality of the leader by the outcome of a policy; nevertheless, the voter's preference captures his general interest in the policy outcome. To see this, suppose a leader is reelected for the second and the last term. Since the leader is no longer concerned about reelection, she will follow her preference in choosing a policy. Specifically, while an unbiased leader will choose a policy that she believes to be the most appropriate, a biased leader will be inclined to choose policy $x = 1$ since it brings her a private benefit. With these last-term incentives in mind, the voter is better off making his voting decision based on his updated belief about the leader's type rather than on current policy outcome, which is a sunk cost. Note that the voter's equilibrium strategy remains simple given the utility function: replace an incumbent only if his updated belief about the leader's type is strictly less than the prior (π_0).

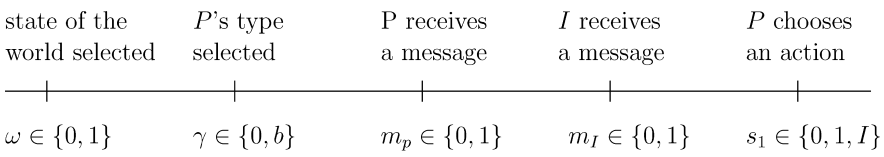
The game begins with nature choosing a state of the world $\omega \in \{0, 1\}$ (Figure 1). The true state of the world is unknown to both P and I, but there is a common prior belief that each state occurs with probability 1/2. Next, P and I receive private messages $m_i \in \{0, 1\}$ ($i = P, I$) regarding the state of the world. Each player's information is independent of the other's given the state of the world and is drawn from a state-dependent distribution satisfying

$$Pr[m_P = 0 \mid \omega = 0] = Pr[m_P = 1 \mid \omega = 1] = \beta_P,$$

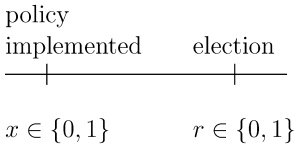
$$Pr[m_I = 0 \mid \omega = 0] = Pr[m_I = 1 \mid \omega = 1] = \beta_I.$$

[9] The utility function of the voter implies that he will vote against a leader *for sure* if the leader is perceived to be biased in her foreign policy conduct. Any one-dimensional model that suggests a democratic leader's electoral fortune be determined by foreign policy alone is obviously too simplistic. It can be proven, however, that all the equilibrium results hold if the model is modified to assume that a leader loses her reelection with a higher probability if the voter does not approve of her performance on foreign policy.

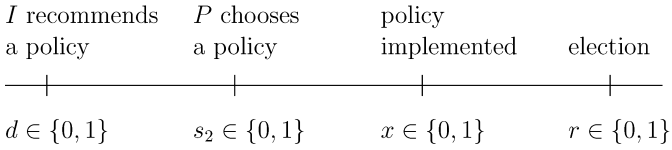
FIGURE 1 Time Line of the Institutional Game



(a) If P chooses a unilateral action $(s_1 = 0, 1)$:



(b) If P chooses to consult the institution $(s_1 = I)$:



Moreover, assume that the private messages are informative, i.e., they improve on players' prior beliefs about the world, and the quality of the information received by I is higher than that received by P. Thus, $1/2 \leq \beta_P < \beta_I \leq 1$.

Nature also chooses a type $\gamma \in \{0, b\}$ for P, and P alone learns the type. P then decides what foreign policy to adopt. If P decides to take a unilateral action, then the chosen policy $x \in \{0, 1\}$ is implemented and the game enters the election stage. On the other hand, if P decides to consult I, then I recommends either $x = 0$ or $x = 1$ based on its private message. After I's recommendation is made public, P decides what her policy is going to be. Note that at this stage P has the option of complying or defying I's decision. The policy is then implemented and the game enters the election stage. Finally, at the election stage v casts the vote.

The strategies of players are mappings specifying actions as functions of the history of the game. A strategy for the leader is a pair of functions s_1 and s_2. Function s_1 maps P's type and P's private message to an action set, $s_1 : \{0, b\} \times \{0, 1\} \to \{0, 1, I\}$, where $s_1 = 0$ and $s_1 = 1$ represent unilateral actions, and $s_1 = I$ represents the action *consult the institution*. If P takes an issue to the institution, then function s_2 maps P's type, P's private message, and I's recommendation to a policy choice, $s_2 : \{0, b\} \times \{0, 1\} \times \{0, 1\} \to \{0, 1\}$. A strategy for the institution, if it is consulted, is a mapping from the message that it receives to its recommendation of a policy, $d : \{0, 1\} \to \{0, 1\}$. Lastly, the voter's strategy is a mapping of history of the game to the vote choice, $r : H_t \to \{0, 1\}$. There are two possible histories of the game, i.e., $H_t = \{\bar{I}\} \times \{0, 1\}$ or $H_t = \{I\} \times \{0, 1\} \times \{0, 1\}$, where $\bar{I}$ represents the leader taking a unilateral action.

The model is a game of incomplete information, and it is solved for pure strategy perfect Bayesian equilibria (PBE). There are multiple pure strategy PBE to this game, some of which are only supported by unreasonable off-equilibrium-path beliefs of the voter.[10] Consequently, I restrict voter's off-equilibrium-path beliefs so that we can focus on substantively interesting equilibria. Specifically, assume that seeing a deviation from the equilibrium path to policy $x = 0$, the voter infers that the leader is more likely to be an unbiased type; seeing a deviation from the equilibrium path to policy $x = 1$, the voter infers that the leader is more likely to be a biased type. The intuition is that since a biased leader favors policy $x = 1$ over

[10]In one such equilibrium the unbiased type takes an issue to the institution but does just the *opposite* of what the institution recommends. This equilibrium can be supported if the voter's off-equilibrium-path belief about any unexpected action by a leader is that the leader is a biased type. The assumption adopted here eliminates this sort of "odd" equilibrium behavior.

$x = 0$, off the equilibrium path, a biased leader is more likely to take on policy $x = 1$ than an unbiased leader, and less likely to take on policy $x = 0$ than an unbiased leader. Additionally, assume that when a unilateral action brings the same payoff as consulting the institution, a leader will choose the unilateral action. In other words, it is assumed that there is a minimal cost, ϵ, involved in consulting the institution.

Equilibrium Analysis

The analysis in this section establishes the claim that international institutions can change the behavior of democratic leaders by influencing domestic politics. Because the study focuses on the informational role of international institutions, the class of equilibria analyzed here are the ones in which the institution's policy recommendation is consistent with the message that it received, i.e., $d = m_I$.[11] The following two lemmas are the building blocks of the main results.[12]

Lemma 1 characterizes the behavior of unbiased leaders after they consult the institution. Recall that in the model a leader does not have to follow what the institution suggests after taking an issue to the institution—a feature that is consistent with the view that international institutions do not have means to enforce their decisions. However, the following lemma states that in equilibrium if an unbiased leader consults the institution, then she will indeed take the institution's advice.

Lemma 1. *In any equilibrium, if* $s_1(0, m_P) = I$, *then* $s_2(0, m_P, d) = d$ *for every* (m_P, d).

Since the institution receives a stronger signal about the state of the world than the leader, an unbiased leader naturally leans toward following what the institution suggests after taking the issue to the institution. What may induce the unbiased type to defy the institution is if the recommended policy causes her to lose reelection. This would be the case if the biased type is more likely to adopt the strategy of following what the institution suggests, so that seeing such a strategy the voter would revise down his prior belief about the leader's type. But then, the unbiased type would be better off not consulting the institution at all and taking the unitary action $x = 0$, because it will signal to the voter that she is an unbiased type and get her reelected. The utility gain from staying in power will be more than enough to compensate for the possible loss in policy outcome for the leader. Therefore, in equilibrium an unbiased leader will only consult the institution if she intends to take the institution's advice.

The next result strengthens Lemma 1 by further characterizing the unbiased type's incentive to consult the institution. While Lemma 1 describes how an unbiased leader will behave *after* taking an issue to the institution, it is agnostic about the circumstance under which she *will* consult the institution. One may wonder if the decision is contingent on the private message received by the leader. Lemma 2 states that it is not. In other words, if an unbiased leader has an incentive to consult the institution having received one message, then the leader is equally motivated to do so after receiving the other message.

Lemma 2. *In any equilibrium, if* $s_1(0, m_P) = I$, *then* $s_1(0, 1 - m_P) = I$.

Because following the institution's advice results in a higher chance of choosing a correct policy, it seems straightforward for an unbiased leader to always seek advice from the institution. However, electoral concerns may interfere with this incentive since getting reelected brings a higher payoff than the expected utility for choosing a correct policy. Therefore, we need to rule out the possibility that the leader consults the institution only when receiving a particular message. Consider the scenario in which an unbiased leader seeks the institution's advice when $m_P = 0$, but takes a unilateral action when $m_P = 1$. This can only be an equilibrium strategy if taking the unilateral action gets the leader reelected while consulting the institution does not. If this is the case, however, the strategy of going to the institution when $m_P = 0$ cannot be part of an equilibrium strategy; the leader has an incentive to deviate to the unilateral action. Lemma 2 thus provides an important characterization of unbiased leaders' behavior in equilibrium: they either always go to the institution, or they never go to the institution. Furthermore, by Lemma 1 once unbiased leaders consult the institution they will follow the institution's recommendation.

The two lemmas combined pin down some aspects of the leader's equilibrium strategy that involves consulting the institution. Based on these results, the first proposition states the nonexistence of an important class of separating equilibria.

[11] Technically, the mapping from I's message to its recommendation can take two other forms in pure strategy PBE. One scenario is that the institution recommends the policy that contradicts the information that it receives: $d = 0$ if $m_I = 1$, and $d = 1$ if $m_I = 0$. Because the relationship between d and m_I is one-on-one in this scenario, the leader can *back out* what m_I is from d, thus the equilibrium results do not change. The second scenario is that the institution's recommendation is not contingent on its information: $d = 0$ (or $d = 1$) for either message. Then the institution's decision is uninformative, and there is no reason for the leader to consult the institution.

[12] All proofs are found in the appendix.

Proposition 1. *There is no fully separating equilibrium in which one type of leader always consults the institution and the other type only takes unilateral actions.*

The implication of the proposition is that if there exists an equilibrium in which one type of leader consults the institution, then the other type will also consult the institution in the equilibrium. In particular, if unbiased leaders' equilibrium strategy is to consult the institution, then a biased leader is forced to behave similarly because a unilateral action will reveal her type to the voter and cause her an electoral loss. On the other hand, if the unbiased type takes a unilateral action, then the biased type has no incentive to consult the institution alone. The two often conflicting incentives of biased leaders—their electoral concerns and their private agendas—work to eliminate the possibility of full separation of behavior. The result highlights a powerful signaling effect of international institutions. The leader in the model has a choice of not consulting the institution and not letting "undesirable" information to be revealed to the voter, yet a biased leader cannot circumvent the institution if the voter believes that an unbiased leader would consult the institution.

Proposition 1 states that the biased type mimics the unbiased type in taking an issue to the institution, but it does not specify whether she will follow the institution's advice later on. From Lemma 1 we know that once an issue is brought to the institution, the unbiased type will adopt the recommended policy; however, it is possible that the biased type will behave differently at this stage. The biased type may pool with the unbiased type as far as she could because of electoral concerns, but separate in the end if the size of the private benefit from policy $x = 1$ is sufficiently large. The next two propositions specify when a biased leader will act *exactly* like the unbiased type and when she will reveal her "true color" to the voter.

To facilitate the statement of the propositions, I first introduce some shorthands. Define a *unilateral equilibrium* as an equilibrium in which only unilateral actions are taken by both types, and an *institutional equilibrium* as one that involves a consultation of the institution by at least one type. For the latter category, define a *fully pooling institutional equilibrium* as an equilibrium in which both types consult the institution given either message and subsequently follow what the institution recommends. In contrast, define a *partial pooling institutional equilibrium* as one in which the biased type mimics the behavior of the unbiased type in consulting the institution given either message, but then ignores the institution's advice sometimes. Which institutional equilibrium will emerge depends on the size of the private benefit derived from policy $x = 1$.

Proposition 2. *There is a unique fully pooling institutional equilibrium if and only if* $0 \leq b < 1 + \frac{\alpha(\beta_I - \beta_P)}{\beta_P + \beta_I - 2\beta_P\beta_I}$.

This is an equilibrium in which the biased type fully mimics the "good" behavior of the unbiased type and follows the institution's advice.[13] Let $B = 1 + \frac{\alpha(\beta_I - \beta_P)}{\beta_P + \beta_I - 2\beta_P\beta_I}$. Note that $B > 1$, i.e., the upper bound of the private benefit for the equilibrium to exist is greater than the electoral benefit. This means that the biased type is content to behave like the unbiased type even when the private benefit is relatively large. Biased leaders do not have an incentive to deviate to a unilateral action when the institution recommends the policy that they prefer, i.e., $x = 1$; however, they may consider a deviation when the recommended policy is $x = 0$. If they do deviate in that case, they will not only lose the reelection by revealing their type, but also forgo a better policy outcome from following the institution's advice. Therefore, it takes more than compensating for the electoral loss to induce the biased type to defy the institution in equilibrium. Exactly how much more is needed depends on the leader's private message. It is more difficult to prevent a biased leader from deviating when $m_P = 1$ than when $m_P = 0$ because in such a case the chance that policy $x = 1$ happens to be the correct policy is higher. Proposition 2 gives the condition that prevents a deviation in this scenario, which is stricter than what is needed to prevent a deviation when $m_P = 0$.

Several comparative statics can be derived from Proposition 2. First, $\partial B/\partial\alpha > 0$, which means that as the probability for a correct policy to succeed increases, a bigger private benefit is necessary to induce the biased type to defy the institution. As a result, it is easier to sustain the fully pooling behavior when α is large. Second, $\partial B/\partial\beta_P < 0$, i.e., as the quality of the leader's information gets better, it takes a smaller amount of private benefit for the biased leader to deviate from the institution's advice. The reason is that as her own message becomes more reliable, a biased leader is more likely to adopt policy $x = 1$ when the message contradicts what the institution suggests. Finally, $\partial B/\partial\beta_I > 0$, which means that the threshold increases as the quality of the information received by the institution gets better, and it is easier to prevent biased leaders from ignoring the institution's advice. The last two points are also captured by the positive relationship between $\beta_I - \beta_P$ and B. The better the information received by the institution relative to that received by the leader, the stronger the incentive to utilize the institution, and the larger the private benefit has to be to induce the biased type to

[13] All the uniqueness results in the article refer to the uniqueness of the specific type of equilibrium characterized in each proposition. Proposition 4 offers a condition under which a general uniqueness result can be obtained.

deviate. Alternatively, as the quality of the leader's private information approaches that of the institution, there is less incentive to go to the institution for an endorsement.

The next proposition shows that when the private benefit is sufficiently large, biased leaders only pool partially with unbiased leaders in consulting the institution. There are two partial pooling institutional equilibria to this game, which are characterized by part (i) and (ii) in Proposition 3. Part (i) characterizes an equilibrium in which the biased type always ignores the institution's advice when it recommends policy $x = 0$; part (ii) characterizes the other partial pooling institutional equilibrium in which the biased type defies the institution only when her private message supports policy $x = 1$.

Proposition 3. *(i) If and only if* $b > 1 + \frac{\alpha(\beta_P+\beta_I-1)}{1-\beta_P-\beta_I+2\beta_P\beta_I}$, *there is a unique partial pooling institutional equilibrium in which the biased type defies the institution whenever the institution recommends policy* $x = 0$. *(ii) If and only if* $1 + \frac{\alpha(\beta_I-\beta_P)}{\beta_P+\beta_I-2\beta_P\beta_I} < b < 1 + \frac{\alpha(\beta_P+\beta_I-1)}{1-\beta_P-\beta_I+2\beta_P\beta_I}$, *there is a unique partial pooling institutional equilibrium in which the biased type defies the institution only if her private message is* $m_P = 1$ *when the institution recommends policy* $x = 0$.[14]

Obviously, in a partial pooling institutional equilibrium there is some chance that a biased leader will have her type revealed and lose reelection. For a biased leader to play such a strategy in equilibrium, the private benefit has to be greater than 1, the value of staying in power. Additionally, the condition in (i) is stricter than that in (ii) since it is more costly for biased leaders to deviate *whenever* the institution recommends $x = 0$ than if the deviation is contingent on m_P.

Let $B' = 1 + \frac{\alpha(\beta_P+\beta_I-1)}{1-\beta_P-\beta_I+2\beta_P\beta_I}$. It can be shown that $\partial B'/\partial\beta_I > 0$, and $\partial B'/\partial\beta_P > 0$. As the quality of the institution's information improves, there is more incentive to follow its recommendation for better policy outcome, and a larger private benefit is necessary to induce a biased leader to deviate. Similarly, as the quality of the leader's private information gets better, it is harder to deviate to $x = 1$ when both her private message and the institution's message point to $x = 0$ being the appropriate policy, thus more compensation is required for the deviation. When $B < b < B'$, the biased type is sensitive to her own message and will only defy the institution if $m_P = 1$. Propositions 2 and 3 combined give a complete characterization of the institutional equilibria: for any value of b, there is a unique institutional equilibrium identified by one of the two propositions.

In light of these results, a natural question is this: are there any unilateral equilibria to this game, in which neither type consults the institution? While the equilibrium analysis so far has focused on institutional equilibria, the model does not rule out the possibility that leaders take actions without consulting an international institution. In other words, unilateral equilibria may coexist with an institutional equilibrium for a parameter range. The existence of unilateral equilibria, however, depends on the voter's off-equilibrium-path belief about a leader's type when the voter observes an unexpected consultation of the institution by the leader. Specifically, if the voter associates with such behavior a belief that the leader is more likely a biased type, then unilateral equilibria can exist. On the other hand, if seeing such behavior the voter maintains his prior belief about the leader's type or revises it upward, then no unilateral equilibrium can be supported. In such a case, an unbiased leader will always have an incentive to deviate from a unilateral action and consult the institution, because the deviation will give her a better policy outcome without reducing her chance of getting reelected. With this additional refinement of the voter's off-equilibrium-path belief, fully pooling and partial pooling institutional equilibria characterized in Propositions 2 and 3 are the only possible equilibria to the game. The next proposition summarizes the result.

Proposition 4. *If the voter's belief about a leader's type does not decrease when the voter observes an unexpected consultation of the institution by the leader, then an institutional equilibrium is the only equilibrium to the game.*

The voter may have an off-equilibrium-path belief described in Proposition 4 if he values an endorsement by an international institution so that his assessment of a leader's type improves when the leader consults the institution. The voter's belief in turn eliminates the incentive to take unilateral actions for both types of leaders. In particular, the biased type is compelled to consult the institution to take a chance at an endorsement, while keeping the option of defying the institution on the table.

Returning to the case of the Second Gulf War, it can be argued that the course President Bush took to go to war with Iraq was influenced by American public opinion, and his behavior resembled that predicted by partial pooling institutional equilibrium in Proposition 3 (ii). Public opinion polls taken in the months leading up to the war consistently showed that a majority of Americans favored the United States acquiring UN approval before a military action (Kull 2002a, 2002b; PIPA/Knowledge Networks 2003a, 2003b). Strikingly, the

[14]The uniqueness in Propositions 2 and 3 are for the strict inequalities. At the two cutpoints, i.e., $b = 1 + \frac{\alpha(\beta_I-\beta_P)}{\beta_P+\beta_I-2\beta_P\beta_I}$ and $b = 1 + \frac{\alpha(\beta_P+\beta_I-1)}{1-\beta_P-\beta_I+2\beta_P\beta_I}$, there are two equilibria—the equilibrium to the left of the cutpoint and the equilibrium to the right of the cutpoint.

public made the distinction between a UN-authorized action and an ally-supported action, and a majority favored the former (PIPA/Knowledge Networks 2003b). It seemed that going around the UN altogether and taking a unilateral (or even an allied) action was not much of an option for Bush as late as mid-February 2003. The British Prime Minister, Tony Blair, Bush's strongest ally, faced a similar situation at home and urged Bush to ask for a second UN resolution on Iraq,[15] which Bush obliged in late February (Woodward 2004, 297). The "danger" of consulting an international institution, however, is that the institution may not recommend the policy that a leader prefers. It was clear by March 12, after Bush called the presidents of Mexico and Chile and received negative responses, that the second resolution would not pass the UN Security Council (Woodward 2004, 343–45). A few days later, Bush withdrew the resolution and went ahead with the war. It is plausible that Bush had had a preference for a military solution to the Iraq problem, and when his private information contradicted UN inspectors' reports and supported his preferred policy, he chose to implement the policy without a UN authorization.

The Welfare-Enhancing Effect of the Institution

To fully understand the implications of the results in the last section, it is helpful to compare the institutional equilibria with the unilateral equilibria from a hypothetical world without an international institution. For the comparison, the original model is modified to allow for only unilateral actions, and the next two propositions are based on the equilibrium analysis of the modified model.[16]

Proposition 5. *In a world without an international institution, if* $b \geq 1 + \alpha(2\beta_P - 1)$, *there exists a unique fully separating unilateral equilibrium. In the equilibrium, the unbiased type always chooses policy* $x = 0$, *and the biased type always chooses policy* $x = 1$.

Generally speaking, from the voter's welfare perspective, full separation in leaders' behavior is not desirable since it implies that biased leaders choose different policies from unbiased leaders.[17] By Proposition 1, in a world with an international institution, as long as there exist unbiased leaders who will consult the institution for a better policy outcome, fully separating behavior will not be observed; biased leaders will be deterred from freely adopting their preferred policy and mimic unbiased leaders to some degree. Proposition 5 says this is no longer the case without such an institution. Furthermore, in the fully separating unilateral equilibrium, even an unbiased leader does not choose a policy contingent on her message; she has to adopt policy $x = 0$ in order to signal her type to the voter. This leads to a welfare loss for the voter. In contrast, the existence of an international institution eases the signaling problem for an unbiased leader, and she can adopt the policy that is more likely to succeed. Finally, the condition for the fully separating unilateral equilibrium to exist, $b \geq 1 + \alpha(2\beta_P - 1)$, is less binding than the condition for the partial pooling institutional equilibrium in Proposition 3 (i) to exist. This means that undesirable separating behavior is more likely to occur without an international institution.

The next proposition examines the condition for pooling behavior to exist in the hypothetical world.

Proposition 6. *In a world without an international institution, if* $b \leq \alpha(2\beta_P - 1)$, *there exists a fully pooling unilateral equilibrium in which both types of leaders choose policies that are consistent with the messages.*

The condition for this intuitively appealing fully pooling unilateral equilibrium to exist is more restrictive than that in Proposition 2, which means that it is more difficult to sustain the beneficial pooling behavior without the institution.[18]

Propositions 5 and 6 highlight two aspects of the informational role of international institutions. First, since international institutions could provide better information about the world, the opportunity costs of ignoring

[15]The first resolution refers to UN Resolution 1441, which urged Iraq to disarm or face "serious consequences."

[16]The proofs are available upon request.

[17]There is arguably a welfare gain for the voter if different types of leaders behave differently—the voter makes fewer reelection mistakes, which leads to better future policy outcomes. So how do the two scenarios compare given this additional consideration? In a world without an international institution, the voter could get a bad policy outcome in the current period, but will be able to replace a biased leader after learning her type. The probability that the new leader is an unbiased type is π_0. In a world with an international institution, if different types of leaders behave differently in an institutional equilibrium, then the voter could also keep an unbiased leader in office and throw out a biased leader. On the other hand, if the biased type fully mimics the equilibrium behavior of the unbiased type, then the voter gets a better policy outcome in the first period and reelects a leader who is unbiased with probability π_0. If the new leader turns out to be biased, then the voter gets a bad policy outcome in the second period. Assuming that the voter discounts future payoffs to some degree, the voter is better off having a biased leader who behaves like an unbiased type in the current period.

[18]There is another fully pooling equilibrium in which both types of leaders always take on policy $x = 0$. The condition for this equilibrium to exist is $b \leq 1 + \alpha(1 - 2\beta_P)$.

their recommendations and not doing the right thing are higher for leaders when such institutions exist. Consequently, larger compensations are necessary to induce biased leaders to deviate from an appropriate policy, and so deviations occur less frequently in the presence of international institutions. Secondly, an international institution could serve as a signaling device about a leader's type by suggesting what the appropriate policy is. When such a signaling device is not available, an unbiased leader has a more difficult time separating herself from the biased type and could be forced to choose a policy that is not in the best interest of the voter.

Public Opinion and Multilateralism in U.S. Use of Force

Empirically, leaders sometimes consult international institutions but not always. In addition, they choose to consult a particular institution but not others. To what extent can the model shed light on such empirical variability? While an extensive empirical analysis is beyond the scope of the article, a preliminary examination of the U.S. presidential use of force can nevertheless provide some answers to these questions.

U.S. presidents used military force 212 times from 1948 to 1998 and obtained an authorization from an international institution in 29 of these cases (Tago 2005). This clearly indicates the coexistence of unilateral and institutional equilibria. The distribution of the 29 cases shows an interesting distinction between the Cold War and post–Cold War periods. Out of the 29 cases, 17 occurred during the post–Cold War period, comprising almost a third of the 53 cases of presidential use of force in this period. In contrast, the remaining 12 cases comprise less than 8% of 159 cases of U.S. use of force during the Cold War period. Considering that the United States has become the sole superpower after the Cold War, thus having more freedom to act unilaterally, the pattern is puzzling.

Most recent studies on U.S. public opinion find that there is considerable stability in public opinion on foreign policy issues, and Americans have a strong preference for multilateralism over "going it alone," especially when it comes to use of force (Eichenberg 2005; Holsti 2004; Kull and Destler 1999; Sobel 2001). Such a preference interacts with specific circumstances and may allow leaders to have more or less freedom in choosing unilateral actions. Sobel (2001, 230) finds that the four administrations since Vietnam showed intermittent reluctance to cooperate with UN efforts, but public attitudes lessened the likelihood of unilateral military actions because "more cooperative policies tended to have stronger support from the American people." In line with Proposition 4, a plausible explanation of the observed pattern in U.S. use of force is that compared with the post–Cold War era, in which the United States has been the dominant power, the American public was more willing to give benefit of the doubt to presidents for unilateral actions during the Cold War when there was a significant security threat from the Soviet Union. After the Cold War, however, with the threat to the nation's survival removed, presidents are subject to greater influence of public opinion, which has been consistently in favor of multilateralism. This is perhaps most clearly demonstrated in the Bosnia crisis.

A series of polls taken between December 1992 and May 1993 showed that there was a majority support for U.S. intervention in Bosnia *only if* it was done as part of multilateral operation through the UN or NATO (Sobel 2000). After abandoning his preferred policy, "lift and strike," a potential U.S. unilateral military action aiming at arming Muslim forces in Bosnia, President Clinton was slow in committing to alternative military actions and held out for political solutions for two years. However, as the American public was growing increasingly supportive of air strike and military action on a multilateral basis, Clinton became more forceful in his addresses to the public, and in mid-1995, the United States engaged in an intense NATO military action authorized by the UN (Sobel 2001, 212–18).

Another notable aspect of the multilateral use of force is that there is a variation in the institutions that authorized such actions. The model developed in this study assumes that there is a single international institution for the consultation of a foreign policy issue; empirically, however, there can be more than one relevant institution for leaders to choose from. What kinds of considerations go into a president's decision to consult a particular institution? The UN seems to be the focal institution in the minds of the American public, beating even NATO as the most favored option for leading a multilateral action (Kull and Destler 1999, 79). At the same time, Americans have reservations about the effectiveness of the UN in carrying out a task (Sobel 2000, 118). This skepticism of UN performance may leave room for presidents to use alternative institutions. Moreover, strategic considerations could play a role in the choice of an institution. For example, during the Cuban Missile Crisis, President Kennedy took the issue to the Organization of American States for an authorization instead of the UN Security Council, knowing that the Soviets would veto it (O'Connell 2000). More recently, the United States acquired an authorization from NATO to use force in Kosovo in 1998

when a resolution from the UN Security Council was not forthcoming. In both cases, support from an international institution was solicited at the pressure of the public, but the presidents were able to choose an institution that would approve a policy that they prefer. This is an important aspect of the strategic interaction between state leaders and international institutions that is not addressed in the article and deserves to be investigated in future research.

Conclusion

This article develops a game-theoretic model that provides a microfoundation for the influence of international institutions on state behavior. While a significant body of empirical studies has emerged demonstrating the influence of international institutions on foreign policy issues ranging from international conflict to global environment, little theoretical work exists that specifies the underlying causal mechanisms. In an era when international institutions seek to play an increasingly active role in promoting peace and prosperity, it is important to understand how they could pursue these goals effectively.

The most significant finding of the article is that, combined with democratic elections, an international institution could put constraints on the behavior of leaders with private agendas. The logic behind the result is that voters would like to reelect an unbiased leader and unbiased leaders have an incentive to take a foreign policy issue to an institution for a better policy outcome; their behavior in turn compels biased leaders to behave similarly out of electoral concerns. As long as voters believe that unbiased leaders exist, an institutional equilibrium exists. If voters additionally hold a favorable view about a particular international institution, then we have a stronger result in which *only* an institutional equilibrium exists and voters receive a welfare improvement. How voters view an international institution thus critically influences leaders' foreign policy behavior, and an international institution ought to maintain a desirable public image, perhaps by working directly with domestic groups, in order to play a larger role in world politics.

The model does not intend to explain the world perfectly; it examines the set of cases where positions taken by international institutions *do* convey useful information. Of course, given that states are strategic actors with self-interests, it is naive to believe that they will always truthfully reveal their intentions and beliefs. Nevertheless, it seems reasonable to assume that the institutions are informative in nontrivial circumstances, which is all that is required for the equilibrium results to hold. When such a minimal condition is not satisfied, an international institution has no informational advantage at all and we need to look elsewhere for an explanation of its influence.

Appendix

A.1 Proof of Lemmas 1 and 2

Lemma 1. *In any equilibrium, if* $s_1(0, m_P) = I$, *then* $s_2(0, m_P, d) = d$ *for every* (m_P, d).

Proof. The lemma says that if in any equilibrium the unbiased type ($\gamma = 0$) consults the institution, then she will subsequently adopt the policy recommended by the institution. To prove the lemma, first consider the following claim:

Claim. *In equilibrium, if* $s_1(0, m_P) = I$, *then* $s_2(0, m_P, d)$ *is the same for* $m_p = 0$ *and* $m_P = 1$.

The claim states that on the equilibrium path, if an unbiased leader appeals to I and I gives recommendation d, then the leader has a unique best response that is not contingent on her own message. Suppose P takes issues to I receiving some message m_P. Based on P's equilibrium strategy, I's decision, and P's subsequent action, v forms a belief about P's type and votes accordingly. Given I's decision and v's belief, for type $\gamma = 0$, one of the two strategies, $s_2(0, m_P, d) = 0$ or $s_2(0, m_P, d) = 1$, is strictly better than the other *irrespective* of her private message. To see this, note that since $\alpha \in (0, 1)$ (the probability that a correct policy succeeds), the expected utility for P from policy outcome is strictly less than 1; on the other hand, being reelected gives her a utility of 1. Therefore, if there is a conflict between an action leading to a better policy outcome and an action leading to her reelection, type $\gamma = 0$ will choose the latter without considering her own message. If both actions lead to the same election result, then she will choose the one that gives her a higher probability of policy success, which is the action recommended by I. Thus, given a recommendation by I, type $\gamma = 0$ has a best response that is not contingent on her private message. This proves the claim.

The claim allows us to look at the unbiased type's equilibrium strategy after I's decision without referring to the specific message she receives. Now we proceed to prove the lemma.

First, note that abiding by I's recommendation always results in a higher probability of policy success for either type of P. What may induce type $\gamma = 0$ to adopt a different policy from I's recommendation is her concern for the

voter's belief, which determines her reelection chance. The voter's belief in turn depends on what type $\gamma = b$ does in the equilibrium. In what follows I show that if type $\gamma = 0$ consults the institution in equilibrium, then she has no incentive to deviate from what I recommends.

(1) Suppose in equilibrium type $\gamma = 0$ consults the institution given a message m_P and then takes action 0 if the institution recommends action 0. That is, $\gamma = 0$ plays the following equilibrium strategy: $s_1(0, m_P) = I$, and $s_2(0, m_P, 0) = 0$. As noted above, the strategy gives the leader the best policy outcome given that the institution receives a stronger signal than the leader. Now suppose the strategy also gets $\gamma = 0$ reelected. Then the leader receives the highest total payoff possible and thus has no incentive to deviate. Next, suppose the strategy does not get the leader reelected. This means that the *biased* type is more likely to play the strategy so that seeing the strategy, the voter revises down his belief about the leader's type from the prior. This can only happen if the biased type plays $s_1(b, m_P) = I$ and $s_2(b, m_P, 0) = 0$ receiving *either* message. But then the unbiased leader would be better off not consulting the institution in the first place and taking the unilateral action $s_1(0, m_P) = 0$, because it will get her reelected and thus bring a higher payoff. A contradiction.

(2) Suppose in equilibrium type $\gamma = 0$ consults the institution given a message m_P and then takes action 1 if the institution recommends action 1. That is, $\gamma = 0$ plays the following equilibrium strategy: $s_1(0, m_P) = I$, and $s_2(0, m_P, 1) = 1$. Again, the strategy gives the leader the best policy outcome. Now suppose the strategy also gets $\gamma = 0$ reelected. Then the leader receives the highest total payoff possible and thus has no incentive to deviate. Next, suppose the strategy does not get the leader reelected. This means that the *biased* type is more likely to play the strategy so that seeing the strategy, the voter revises down his belief about the leader's type from the prior. This can only happen if the biased type plays $s_1(b, m_P) = I$ and $s_2(b, m_P, 1) = 1$ receiving *either* message. But then the unbiased leader would be better off not consulting the institution in the first place and taking the unilateral action $s_1(0, m_P) = 0$, because it will get her reelected and the utility from holding office will be more than enough to compensate for the loss in policy outcome by taking the unilateral action. A contradiction.

In sum, taking an action that is consistent with I's recommendation after seeking I's advice is part of the equilibrium strategy for $\gamma = 0$. That is, if $s_1(0, m_P) = I$, then $s_2(0, m_P, d) = d$ for every pair of (m_P, d). □

Lemma 2. *In any equilibrium, if* $s_1(0, m_P) = I$, *then* $s_1(0, 1 - m_P) = I$.

Proof. The lemma says that if the unbiased type has an incentive to consult the institution when receiving one message, then the type is equally induced to do so receiving the other message. First, we calculate the expected payoff from policy outcome when $\gamma = 0$ consults I. By Lemma 1, if $s_1(0, m_P) = I$, then $s_2^*(0, m_P, d) = d$. Suppose $m_P = 0$. Then, P's belief that $\omega = 0$ is

$$\begin{aligned}\mu'_P &= Pr(\omega = 0 \mid m_P = 0)\\ &= \frac{1/2\beta_P}{1/2\beta_P + 1/2(1-\beta_P)} = \beta_P\end{aligned}$$

If P takes an issue to I, then from P's perspective the probabilities that I receives each message are

$$\begin{aligned}Pr(m_I = 0 \mid m_P = 0) &= \mu'_P\beta_I + \big(1-\mu'_P\big)(1-\beta_I)\\ &= 1 - \beta_P - \beta_I + 2\beta_P\beta_I\\ Pr(m_I = 1 \mid m_P = 0) &= \beta_P + \beta_I - 2\beta_P\beta_I\end{aligned}$$

Correspondingly, $d = 0$ and $d = 1$ with these probabilities. The probabilities that I's recommendation matches the state of the world are $Pr(\omega = 0 | m_P = 0, m_I = 0) = \frac{\beta_P\beta_I}{1-\beta_P-\beta_I+2\beta_P\beta_I}$, and $Pr(\omega = 1 | m_P = 0, m_I = 1) = \frac{(1-\beta_P)\beta_I}{\beta_P+\beta_I-2\beta_P\beta_I}$. Then, $\gamma = 0$'s expected utility from the policy outcome by consulting I is $\alpha\beta_P\beta_I + \alpha(1-\beta_P)\beta_I = \alpha\beta_I$. Similarly, if $m_P = 1$, $\gamma = 0$'s payoff from policy outcome by consulting I is $\alpha\beta_I$.

Now suppose $\gamma = 0$'s equilibrium strategy involves the following: $s_1(0, 0) = I$ and $s_2^*(0, 0, d) = d$, but $s_1(0, 1) = 1$ (or $s_1(0, 1) = 0$). In other words, $\gamma = 0$ consults I if $m_p = 0$, but takes a unilateral action if $m_P = 1$.

Suppose $\gamma = 0$ gets reelected by following I's advice when $m_P = 0$. Now suppose instead of playing the unilateral action $s_1(0, 1) = 1$ when $m_P = 1, \gamma = 0$ deviates to the strategy $s_1(0, 1) = I$ and $s_2(0, 1, d) = d$. Then, she will get reelected and her total expected utility is $EU_P(s_1 = I, s_2 = d \mid m_P = 1) = \alpha\beta_I + 1$,[19] which is better than the most she can get from the unilateral action, $EU_P(s_1 = 1 \mid m_P = 1) = \alpha\beta_P + 1$. Thus, it is a profitable deviation to consult I when $m_P = 1$. A contradiction. Suppose $\gamma = 0$'s strategy of following I's advice when $m_P = 0$ gets her reelected only when $d = 0$, but not when $d = 1$. This can only happen if $\gamma = b$ consults I *and* plays $s_2(b, m_P, d) = 1$ given either massage. This gives $\gamma = b$ a payoff of $\alpha\beta_P + b$. But then, $\gamma = b$ has a profitable deviation to the unilateral action $s_1(b, m_P) = 1$ since it will give her $\alpha\beta_P + b + 1$ (note that she gets reelected by pooling with $\gamma = 0$). A contradiction. Suppose $\gamma = 0$'s strategy of following I's advice when $m_P = 0$ gets her reelected only when $d = 1$, but not when $d = 0$. This can only happen if $\gamma = b$ consults the institution and plays $s_2(b, m_P, d) = 0$

[19] For notational convenience, a strategy $s_i(\cdot)$ is sometimes suppressed as simply s_i.

given either massage. But then, $\gamma = b$ has an incentive to deviate to $s_1(b, m_P) = 1$ since it will give her $\alpha\beta_P + b + 1$. A contradiction. Finally, suppose $\gamma = 0$'s strategy of following I's advice when $m_P = 0$ does not get her reelected at all. This can only happen if type $\gamma = b$ plays the exact same strategy given either value of m_P, so that the voter revises down his belief after observing the strategy. But then, $\gamma = 0$ can profitably deviate to the unilateral action $s_1(0, 0) = 0$, getting reelected and receiving $EU_P(s_1 = 0 \mid m_P = 0) = \alpha\beta_P + 1 > \alpha\beta_I$. A contradiction.

By a similar proof it can be shown that $s_1 = I$ only if $m_p = 1$ cannot be $\gamma = 0$'s equilibrium strategy either. Therefore, if $s_1(0, m_P) = I$, then $s_1(0, 1 - m_P) = I$. □

A.2 Proof of Propositions 1 through 4

Proposition 1. *There is no fully separating equilibrium in which one type of leader always consults the institution and the other type only takes unilateral actions.*

Proof. I prove by contradiction.

(1) First, consider an equilibrium in which type $\gamma = 0$ consults I given either message, while type $\gamma = b$ takes a unilateral action given either message. By Lemma 1, we know that if $\gamma = 0$ consults I, then it will follow I's recommendation. In other words, $\gamma = 0$'s equilibrium strategy is $s_1(0, m_P) = I$ and $s_2(0, m_P, d) = d$. For $\gamma = b$, taking a unilateral action means $s_1(b, m_P) = 0$, or $s_1(b, m_P) = 1$. Because of the separation of the strategies, type $\gamma = b$ does not get reelected in such an equilibrium while type $\gamma = 0$ does. Below I show that no matter what unilateral action $\gamma = b$ takes in the equilibrium, she always has an incentive to deviate.

Case 1: Suppose in the equilibrium type $\gamma = b$ plays $s_1(b, 0) = 0$. In other words, she takes unilateral action 0 given message 0.[20] Her belief that the state of the world is 0 is $\mu'_P(\omega = 0 \mid m_P = 0) = \beta_P$. Her expected utility from taking action $s_1(b, 0) = 0$ is $EU_P(s_1 = 0 \mid m_P = 0) = \alpha\beta_P$. That is, the unilateral action that matches her message gives her a total payoff of $\alpha\beta_P$. Consider a deviation that mimics type $\gamma = 0$'s strategy, that is, $s_1(b, m_P) = I$ and $s_2(b, m_P, d) = d$. The deviation will give her a better policy outcome by following the institution's advice and will get her reelected. She also extracts private benefits whenever the institution recommends policy $x = 1$. Her expected utility from the deviation is $EU_P(s_1 = I, s_2 = d \mid m_P = 0) = \alpha\beta_I + 1 + b(\beta_P + \beta_I - 2\beta_P\beta_I)$. This is a profitable deviation. A contradiction.

Case 2: Suppose in the equilibrium type $\gamma = b$ plays $s_1(b, 0) = 1$.[21] Her belief that the state of the world is 1 is $\mu'_P(\omega = 1 \mid m_P = 0) = 1 - \beta_P$. Her expected utility from taking action $s_1(b, 0) = 1$ is $EU_P(s_1 = 1 \mid m_P = 0) = \alpha(1 - \beta_P) + b$. Consider a deviation that partially mimics type $\gamma = 0$'s strategy: $s_1(b, m_P) = I$ and $s_2(b, m_P, d) = 1$, that is, she consults the institution first, but adopts policy $x = 1$ regardless of what the institution recommends. It gives her the expected utility $EU_P(s_1 = I, s_2 = 1 \mid m_P = 0) = \alpha(1 - \beta_P) + b + (\beta_P + \beta_I - 2\beta_P\beta_I)$. It is a profitable deviation since, in addition to receiving the same payoff as taking the unilateral action, she also wins reelection with probability $\beta_P + \beta_I - 2\beta_P\beta_I$ by pooling with $\gamma = 0$ whenever I recommends $d = 1$. A contradiction.

In sum, there is no fully separating equilibrium in which only the unbiased type consults the institution.

(2) Next, consider an equilibrium in which type $\gamma = b$ consults I given either message, while type $\gamma = 0$ takes a unilateral action given either message. Again, because of the separation of the strategies, type $\gamma = b$ does not get reelected in such an equilibrium while type $\gamma = 0$ does. Below I show that no matter what unilateral action $\gamma = 0$ takes in the equilibrium, type $\gamma = b$ has an incentive to deviate. I first prove a claim which states that after consulting I, type $\gamma = b$ always follows I's advice in this kind of equilibrium.

Claim. *In an equilibrium in which only type $\gamma = b$ consults the institution, $s_2(b, m_P, d) = d$.*

Suppose not. Suppose $\gamma = b$ follows a strategy that leads to $s_2(b, m_P, 1) = 0$ for some m_P. That is, sometimes she will choose policy $x = 0$ when $d = 1$. It is easy to see that deviating to $x = 1$, i.e., following I's recommendation, is profitable in this case since the deviation will give her a better policy outcome *and* a private benefit. A contradiction. So it must be the case that $s_2(b, m_P, 1) = 1$. Suppose $\gamma = b$ follows a strategy that leads to $s_2(b, m_P, 0) = 1$ for some m_P. That is, she will choose policy $x = 1$ when $d = 0$. There are two cases to consider. (1) Suppose $s_2(b, m_P, 0) = 1$ for $m_P = 0$. Combined with the first half of the proof, we know that her expected utility from this strategy when $m_P = 0$ is $\alpha(1 - \beta_P) + b$. But deviating to $s_1(b, 0) = 1$ gives her at least $\alpha(1 - \beta_P) + b$ (i.e., without winning reelection). Since appealing to I involves a small cost by assumption, it is profitable to deviate to the unilateral action. A contradiction. (2) Suppose $s_2(b, m_P, 0) = 1$ for $m_P = 1$. Her expected utility from this strategy when $m_P = 1$ is $\alpha\beta_P + b$. Again, deviating to $s_1(b, 1) = 1$ gives her at least the same payoff, and given that there is

[20] There is no restriction on what she will play when her message is 1: she could play either $s_1(b, 1) = 0$ or $s_1(b, 1) = 1$.

[21] Again, there is no restriction on what she will play when her message is 1: she could play either $s_1(b, 1) = 0$ or $s_1(b, 1) = 1$.

a small cost involved in appealing to I, it is profitable to deviate to the unilateral action. A contradiction.

In sum, for type $\gamma = b$, if $s_1(b, m_P) = I$, then $s_2(b, m_P, d) = d$. This proves the claim.

Now turn to the proposition.

Case 1: Suppose in the equilibrium $\gamma = 0$ plays $s_1(0, m_P) = 1$ for some $m_P \in \{0, 1\}$, and $\gamma = b$ plays $s_1(b, m_P) = I$ and $s_2(b, m_P, d) = d$ given either $m_P \in \{0, 1\}$. If $\gamma = b$ deviates to unilateral action $s_1(b, m_P) = 1$, at worst she gets $EU_P(s_1 = 1 \mid m_P = 0) = \alpha(1 - \beta_P) + b + 1$. That is, in addition to the policy benefit in the worst-case scenario where her policy does not match her message, she will be reelected by pooling with the unbiased type and also receive a private benefit by the strategy. On the other hand, playing her equilibrium strategy, at best she gets $EU_P(s_1 = I, s_2 = d \mid m_P = 1) = \alpha\beta_I + b(1 - \beta_P - \beta_I + 2\beta_P\beta_I)$. So the deviation is profitable. A contradiction. This leaves us with just one more scenario to examine, where $\gamma = 0$ plays $s_1(0, m_P) = 0$ given either $m_P = 0$ or $m_P = 1$ in equilibrium.

Case 2: Suppose in the equilibrium $\gamma = 0$ plays $s_1(0, m_P) = 0$ given either $m_P \in \{0, 1\}$, and $\gamma = b$ plays $s_1(b, m_P) = I$ and $s_2(b, m_P, d) = d$ given either $m_P \in \{0, 1\}$. Suppose $\gamma = b$ receives message $m_P = 0$. Her expected utility from playing her equilibrium strategy is $EU_P(s_1 = I, s_2 = d \mid m_P = 0) = \alpha\beta_I + b(\beta_P + \beta_I - 2\beta_P\beta_I)$. Her expected utilities from deviating to unilateral actions are $EU_P(s_1 = 0 \mid m_P = 0) = \alpha\beta_P + 1$, and $EU_P(s_1 = 1 \mid m_P = 0) = \alpha(1 - \beta_P) + b$. To prevent a deviation, it must be the case that $EU_P(s_1 = I, s_2 = d \mid m_P = 0) \geq EU_P(s_1 = 0 \mid m_P = 0)$, which implies $b \geq \frac{1+\alpha(\beta_P-\beta_I)}{\beta_P+\beta_I-2\beta_P\beta_I}$, as well as $EU_P(s_1 = I, s_2 = d \mid m_P = 0) \geq EU_P(s_1 = 1 \mid m_P = 0)$, which implies $b \leq \frac{\alpha(\beta_P+\beta_I-1)}{1-\beta_P-\beta_I+2\beta_P\beta_I}$. Since $\frac{\alpha(\beta_P+\beta_I-1)}{1-\beta_P-\beta_I+2\beta_P\beta_I} \leq \frac{1+\alpha(\beta_P-\beta_I)}{\beta_P+\beta_I-2\beta_P\beta_I}$, the two inequalities cannot hold simultaneously. A contradiction.

In sum, there is no fully separating equilibrium in which the biased type consults the institution given either message and the unbiased type takes unilateral actions given either message. This completes the proof of Proposition 1.

Proposition 2. *There is a unique fully pooling institutional equilibrium if and only if* $0 \leq b < 1 + \frac{\alpha(\beta_I-\beta_P)}{\beta_P+\beta_I-2\beta_P\beta_I}$.

Proof. I begin by characterizing the equilibrium strategies of all players.

(1) For P:
$s_1(\gamma, m_P) = I$, and $s_2(\gamma, d, m_P) = d$, where $\gamma \in \{0, 1\}$ and $m_P \in \{0, 1\}$.

(2) For I:
$d = m_I$, where $m_I \in \{0, 1\}$.

(3) For v:
$r(\bar{I}, 0) = 1$, $r(\bar{I}, 1) = 0$;
$r(I, 0, 0) = r(I, 1, 1) = 1$;
$r(I, 1, 0) = 1$, and $r(I, 0, 1) = 0$.

Then, strategies (1)–(3) form an equilibrium.

Now check for deviations. Given that P follows I's advice, it is trivial to show that I's strategy specified above is its best response. For v, suppose his belief that P is type $\gamma = 0$ is μ_v. His expected utilities from electing and not electing P given this belief are $EU_v(r = 0) = \pi_0$ and $EU_v(r = 1) = \mu_v$. $EU_v(r = 1) \geq EU_v(r = 0)$ if $\mu_v \geq \pi_0$. That is, v votes for P if his belief about P's type is greater or equal to his prior. On the equilibrium path, since both types pool fully with each other, the voter retains his prior and always reelects P. If P deviates to an action involving policy $x = 0$, by our assumption of v's off-equilibrium-path belief, $\mu_v > \pi_0$, and the voter should reelect P. If P deviates to an action involving policy $x = 1$, by assumption, $\mu_v < \pi_0$, and the voter should replace P. Thus, the voter's strategy specified above is his best response.

For P, by Lemma 2 and the voter's equilibrium strategy, it is trivial to show that the strategy specified above is a best response for type $\gamma = 0$. What is left to check is whether type $\gamma = b$ has an incentive to deviate. There are two cases to consider.

(i) $\gamma = b$ and $m_P = 0$

Given $m_P = 0$, $\mu'_P = Pr(\omega = 0 \mid m_P = 0) = \beta_P$. If P takes an issue to I, the probabilities that I receives each message are $Pr(m_I = 0 \mid m_P = 0) = 1 - \beta_P - \beta_I + 2\beta_P\beta_I$, and $Pr(m_I = 1 \mid m_P = 0) = \beta_P + \beta_I - 2\beta_P\beta_I$. Correspondingly, $d = 0$ and $d = 1$ with these probabilities. The probabilities that I's recommendation matches the state of the world are $Pr(\omega = 0 | m_P = 0, m_I = 0) = \frac{\beta_P\beta_I}{1-\beta_P-\beta_I+2\beta_P\beta_I}$, and $Pr(\omega = 1 | m_P = 0, m_I = 1) = \frac{(1-\beta_P)\beta_I}{\beta_P+\beta_I-2\beta_P\beta_I}$. Then, the expected utility for P to fully pool with type $\gamma = 0$ is $EU_P(s_1 = I, s_2 = d \mid m_P = 0) = \alpha\beta_I + b(\beta_P + \beta_I - 2\beta_P\beta_I) + 1$. Consider P's expected utilities from deviating to unilateral actions: $EU_P(s_1 = 0 \mid m_P = 0) = \alpha\beta_P + 1$, and $EU_P(s_1 = 1 \mid m_P = 0) = \alpha(1 - \beta_P) + b$. It is easy to see that deviating to $s_1 = 0$ is not profitable. To prevent a deviation to $s_1 = 1$, the following inequality must hold:

$$EU_P(s_1 = I, s_2 = d \mid m_P = 0) \geq EU_P(s_1 = 1 \mid m_P = 0)$$

$$b \leq \frac{1 + \alpha(\beta_P + \beta_I - 1)}{1 - \beta_P - \beta_I + 2\beta_P\beta_I} \quad (1)$$

Now check to see if P has an incentive to deviate after I made its decision. If $d = 0$, then $EU_P(s_2 = 0 \mid d = 0) = \frac{\alpha\beta_P\beta_I}{1-\beta_P-\beta_I+2\beta_P\beta_I} + 1$, and $EU_P(s_2 = 1 \mid d = 0) = \frac{\alpha(1-\beta_P)(1-\beta_I)}{1-\beta_P-\beta_I+2\beta_P\beta_I} + b$. To prevent a deviation, the following inequality must hold:

$$EU_P(s_2 = 0 \mid d = 0) \geq EU_P(s_2 = 1 \mid d = 0)$$

$$b \leq 1 + \frac{\alpha(\beta_P + \beta_I - 1)}{1 - \beta_P - \beta_I + 2\beta_P\beta_I} \quad (2)$$

If $d = 1$, then $EU_P(s_2 = 0 \mid d = 1) = \frac{\alpha(1-\beta_I)\beta_P}{\beta_P(1-\beta_I)+(1-\beta_P)\beta_I} + 1$, and $EU_P(s_2 = 1 \mid d = 1) = \frac{\alpha(1-\beta_P)\beta_I}{\beta_P(1-\beta_I)+(1-\beta_P)\beta_I} + b + 1$. The deviation is not profitable.

(ii) $\gamma = b$ and $m_P = 1$

Given $m_P = 1$, $\mu'_P = Pr(\omega = 0 \mid m_P = 1) = 1 - \beta_P$. If P takes an issue to I, the probabilities that I receives each message are $Pr(m_I = 0 \mid m_P = 1) = \beta_P + \beta_I - 2\beta_P\beta_I$, and $Pr(m_I = 1 \mid m_P = 1) = 1 - \beta_P - \beta_I + 2\beta_P\beta_I$. Correspondingly, $d = 0$ and $d = 1$ with these probabilities. The probabilities that I's recommendation matches the state of the world are $Pr(\omega = 0 \mid m_P = 1, m_I = 0) = \frac{(1-\beta_P)\beta_I}{\beta_P+\beta_I-2\beta_P\beta_I}$, and $Pr(\omega = 1 \mid m_P = 1, m_I = 1) = \frac{\beta_P\beta_I}{1-\beta_P-\beta_I+2\beta_P\beta_I}$. Then, the expected utility for P to fully pool with type $\gamma = 0$ is $EU_P(s_1 = I,\ s_2 = d \mid m_P = 1) = \alpha\beta_I + b(1 - \beta_P - \beta_I + 2\beta_P\beta_I) + 1$. Consider P's expected utilities from deviating to unilateral actions: $EU_P(s_1 = 0 \mid m_P = 1) = \alpha(1 - \beta_P) + 1$, and $EU_P(s_1 = 1 \mid m_P = 1) = \alpha\beta_P + b$. It is easy to see that deviating to $s_1 = 0$ is not profitable. To prevent P from deviating to $s_1 = 1$, the following inequality must hold:

$$EU_P(s_1 = I, s_2 = d \mid m_P = 1) \geq EU_P(s_1 = 1 \mid m_P = 1)$$

$$b \leq \frac{1 + \alpha(\beta_I - \beta_P)}{\beta_P + \beta_I - 2\beta_P\beta_I} \quad (3)$$

Now check to see if P has an incentive to deviate after I made its decision. If $d = 0$, then $EU_P(s_2 = 0 \mid d = 0) = \frac{\alpha(1-\beta_P)\beta_I}{\beta_P+\beta_I-2\beta_P\beta_I} + 1$, and $EU_P(s_2 = 1 \mid d = 0) = \frac{\alpha(1-\beta_I)\beta_P}{\beta_P+\beta_I-2\beta_P\beta_I} + b$. To prevent a deviation, the following inequality must hold:

$$EU_P(s_2 = 0 \mid d = 0) \geq EU_P(s_2 = 1 \mid d = 0)$$

$$b \leq 1 + \frac{\alpha(\beta_I - \beta_P)}{\beta_P + \beta_I - 2\beta_P\beta_I} \quad (4)$$

If $d = 1$, then $EU_P(s_2 = 0 \mid d = 1) = \frac{\alpha(1-\beta_P)(1-\beta_I)}{1-\beta_P-\beta_I+2\beta_P\beta_I} + 1$, and $EU_P(s_2 = 1 \mid d = 1) = \frac{\alpha\beta_P\beta_I}{1-\beta_P-\beta_I+2\beta_P\beta_I} + 1 + b$. The deviation is not profitable.

When inequalities (1)–(4) all hold, we have a fully pooling equilibrium. Since (2) is stricter than (1), and (4) is stricter than (3), it comes down to comparing inequalities (2) and (4). It can be shown that $\min\{1 + \frac{\alpha(\beta_P+\beta_I-1)}{1-\beta_P-\beta_I+2\beta_P\beta_I}, 1 + \frac{\alpha(\beta_I-\beta_P)}{\beta_P+\beta_I-2\beta_P\beta_I}\} = 1 + \frac{\alpha(\beta_I-\beta_P)}{\beta_P+\beta_I-2\beta_P\beta_I}$. So, if $b \leq 1 + \frac{\alpha(\beta_I-\beta_P)}{\beta_P+\beta_I-2\beta_P\beta_I}$, then the equilibrium exists; in addition, if and only if $b < 1 + \frac{\alpha(\beta_I-\beta_P)}{\beta_P+\beta_I-2\beta_P\beta_I}$, the equilibrium is unique. This completes the proof of Proposition 2.

Proposition 3. *(i) If and only if* $b > 1 + \frac{\alpha(\beta_P+\beta_I-1)}{1-\beta_P-\beta_I+2\beta_P\beta_I}$, *there is a unique partial pooling institutional equilibrium in which the biased type defies the institution whenever the institution recommends policy* $x = 0$. *(ii) If and only if* $1 + \frac{\alpha(\beta_I-\beta_P)}{\beta_P+\beta_I-2\beta_P\beta_I} < b < 1 + \frac{\alpha(\beta_P+\beta_I-1)}{1-\beta_P-\beta_I+2\beta_P\beta_I}$, *there is a unique partial pooling institutional equilibrium in which the biased type defies the institution only if her private message is* $m_P = 1$ *when the institution recommends policy* $x = 0$.

Proof. Taking part (ii) first, I begin by fully characterizing the equilibrium strategies of all players.

(1) For P:

$s_1(\gamma, m_P) = I$,

$s_2(0, d, m_P) = d$, $s_2(b, d, 0) = d$, $s_2(b, d, 1) = 1$,

where $\gamma \in \{0, 1\}$ and $m_P \in \{0, 1\}$.

(2) For I:

$d = m_I$, where $m_I \in \{0, 1\}$.

(3) For v:

$r(\bar{I}, 0) = 1$, $r(\bar{I}, 1) = 0$;

$r(I, 0, 0) = r(I, 1, 1) = 1$;

$r(I, 1, 0) = 1$, and $r(I, 0, 1) = 0$.

Then, strategies (1)–(3) form an equilibrium.

Now check for deviations. Given other players' strategies, it is trivial to show that I's strategy is its best response. For v, the strategy specified above is his best response (the proof is similar to that for Proposition 2).

For P, by Lemma 2 and the voter's equilibrium strategy, the strategy specified above is a best response for type $\gamma = 0$. What is left to check is the possibility that type $\gamma = b$ deviates. There are two cases to consider.

(i) $\gamma = b$ and $m_P = 0$

Given $m_P = 0$, $\mu'_P = Pr(\omega = 0 \mid m_P = 0) = \beta_P$. If P takes an issue to I, the probabilities that I receives each message are $Pr(m_I = 0 \mid m_P = 0) = 1 - \beta_P - \beta_I + 2\beta_P\beta_I$, and $Pr(m_I = 1 \mid m_P = 0) = \beta_P + \beta_I - 2\beta_P\beta_I$. Correspondingly, $d = 0$ and $d = 1$ with these probabilities. The probabilities that I's recommendation matches the state of the world are $Pr(\omega = 0 \mid m_P = 0, m_I = 0) = \frac{\beta_P\beta_I}{1-\beta_P-\beta_I+2\beta_P\beta_I}$, and $Pr(\omega = 1 \mid m_P = 0, m_I = 1) = \frac{(1-\beta_P)\beta_I}{\beta_P+\beta_I-2\beta_P\beta_I}$. Then, P's expected utility from playing her equilibrium strategy is $EU_P(s_1 = I, s_2 = d \mid m_P = 0) = \alpha\beta_I + b(\beta_P + \beta_I - 2\beta_P\beta_I) + 1$. Consider P's expected utilities from deviating to unilateral actions: $EU_P(s_1 = 0 \mid m_P = 0) = \alpha\beta_P + 1$, and $EU_P(s_1 = 1 \mid m_P = 0) = \alpha(1 - \beta_P) + b$. We can see that deviating to $s_1 = 0$ is not profitable. To prevent a deviation to $s_1 = 1$, the following

inequality must hold:

$$EU_P(s_1 = I, s_2 = d \mid m_P = 0) \geq EU_P(s_1 = 1 \mid m_P = 0)$$

$$b \leq \frac{1 + \alpha(\beta_P + \beta_I - 1)}{1 - \beta_P - \beta_I + 2\beta_P\beta_I} \quad (5)$$

Now check to see if P has an incentive to deviate after I made its decision. If $d = 0$, then $EU_P(s_2 = 0 \mid d = 0) = \frac{\alpha\beta_P\beta_I}{1-\beta_P-\beta_I+2\beta_P\beta_I} + 1$, and $EU_P(s_2 = 1 \mid d = 0) = \frac{\alpha(1-\beta_P)(1-\beta_I)}{1-\beta_P-\beta_I+2\beta_P\beta_I} + b$. To prevent a deviation, the following inequality must hold:

$$EU_P(s_2 = 1 \mid d = 0) \leq EU_P(s_2 = 0 \mid d = 0)$$

$$b \leq 1 + \frac{\alpha(\beta_P + \beta_I - 1)}{1 - \beta_P - \beta_I + 2\beta_P\beta_I} \quad (6)$$

If $d = 1$, then $EU_P(s_2 = 0 \mid d = 1) = \frac{\alpha(1-\beta_I)\beta_P}{\beta_P+\beta_I-2\beta_P\beta_I} + 1$, and $EU_P(s_2 = 1 \mid d = 1) = \frac{\alpha(1-\beta_P)\beta_I}{\beta_P+\beta_I-2\beta_P\beta_I} + b + 1$. The deviation is not profitable.

(ii) $\gamma = b$ and $m_P = 1$

Given $m_P = 1$, $\mu'_P = Pr(\omega = 0 \mid m_P = 1) = 1 - \beta_P$. If P takes an issue to I, the probabilities that I receives each message are $Pr(m_I = 0 \mid m_P = 1) = \beta_P + \beta_I - 2\beta_P\beta_I$, and $Pr(m_I = 1 \mid m_P = 1) = 1 - \beta_P - \beta_I + 2\beta_P\beta_I$. Correspondingly, $d = 0$ and $d = 1$ with these probabilities. The probabilities that I's recommendations will match the state of the world are $Pr(\omega = 0 \mid m_P = 1, m_I = 0) = \frac{(1-\beta_P)\beta_I}{\beta_P+\beta_I-2\beta_P\beta_I}$, and $Pr(\omega = 1 \mid m_P = 1, m_I = 1) = \frac{\beta_P\beta_I}{1-\beta_P-\beta_I+2\beta_P\beta_I}$. Then, P's expected utility from playing her equilibrium strategy is $EU_P(s_1 = I, s_2 = 1 \mid m_P = 1) = \alpha\beta_P + (1 - \beta_P - \beta_I + 2\beta_P\beta_I) + b$. Consider P's expected utilities from deviating to unilateral actions: $EU_P(s_1 = 0 \mid m_P = 1) = \alpha(1 - \beta_P) + 1$, and $EU_P(s_1 = 1 \mid m_P = 1) = \alpha\beta_P + b$. Deviating to $s_1 = 1$ is not profitable. To prevent a deviation to $s_1 = 0$, the following inequality must hold:

$$EU_P(s_1 = 0 \mid m_P = 1) \leq EU_P(s_1 = I, s_2 = 1 \mid m_P = 1)$$

$$b \geq \alpha(1 - 2\beta_P) + (\beta_P + \beta_I - 2\beta_P\beta_I) \quad (7)$$

Now check to see if P has an incentive to deviate after I's decision is made. If $d = 0$, then $EU_P(s_2 = 0 \mid d = 0) = \frac{\alpha(1-\beta_P)\beta_I}{\beta_P+\beta_I-2\beta_P\beta_I} + 1$, and $EU_P(s_2 = 1 \mid d = 0) = \frac{\alpha(1-\beta_I)\beta_P}{\beta_P+\beta_I-2\beta_P\beta_I} + b$. To prevent deviation, the following inequality must hold:

$$EU_P(s_2 = 0 \mid d = 0) \leq EU_P(s_2 = 1 \mid d = 0)$$

$$b \geq 1 + \frac{\alpha(\beta_I - \beta_P)}{\beta_P + \beta_I - 2\beta_P\beta_I} \quad (8)$$

If $d = 1$, then $EU_P(s_2 = 0 \mid d = 1) = \frac{\alpha(1-\beta_P)(1-\beta_I)}{1-\beta_P-\beta_I+2\beta_P\beta_I} + 1$, and $EU_P(s_2 = 1 \mid d = 1) = \frac{\alpha\beta_P\beta_I}{1-\beta_P-\beta_I+2\beta_P\beta_I} + 1 + b$. The deviation is not profitable.

Thus, if inequalities (5)–(8) all hold, we have a partial pooling institutional equilibrium in which the leader defies the institution only if her private message is $m_P = 1$ and the institution recommends policy $x = 0$. It can be shown that (6) is stricter than (5), and (8) is stricter than (7). So, if $1 + \frac{\alpha(\beta_I-\beta_P)}{\beta_P+\beta_I-2\beta_P\beta_I} \leq b \leq 1 + \frac{\alpha(\beta_P+\beta_I-1)}{1-\beta_P-\beta_I+2\beta_P\beta_I}$, then the equilibrium exists; in addition, if and only if $1 + \frac{\alpha(\beta_I-\beta_P)}{\beta_P+\beta_I-2\beta_P\beta_I} < b < 1 + \frac{\alpha(\beta_P+\beta_I-1)}{1-\beta_P-\beta_I+2\beta_P\beta_I}$, the equilibrium is unique.

Back to part (i), I begin by fully characterizing the equilibrium strategies of all players.

(1) For P:

$s_1(\gamma, m_P) = I$,
$s_2(0, d, m_P) = d$, $s_2(b, d, m_P) = 1$,
where $\gamma \in \{0, 1\}$ and $m_P \in \{0, 1\}$.
(2) For I:
$d = m_I$, where $m_I \in \{0, 1\}$.
(3) For v:
$r(\bar{I}, 0) = 1$, $r(\bar{I}, 1) = 0$;
$r^*(I, 0, 0) = r(I, 1, 1) = 1$;
$r(I, 1, 0) = 1$, and $r(I, 0, 1) = 0$.
Then, strategies (1)–(3) form an equilibrium.

Now check for deviations. Given other players' strategies, it is easy to show that I's strategy is its best response. For v, the strategy specified above is his best response (the proof is similar to that for Proposition 2).

For P, by Lemma 2 and the voter's equilibrium strategy, it is a best response for type $\gamma = 0$. What is left to check is whether type $\gamma = b$ has an incentive to deviate. There are two cases to consider.

(i) $\gamma = b$ and $m_P = 0$

Given $m_P = 0$, $\mu'_P = Pr(\omega = 0 \mid m_P = 0) = \beta_P$. If P takes an issue to I, the probabilities that I receives each message are $Pr(m_I = 0 \mid m_P = 0) = 1 - \beta_P - \beta_I + 2\beta_P\beta_I$, and $Pr(m_I = 1 \mid m_P = 0) = \beta_P + \beta_I - 2\beta_P\beta_I$. Correspondingly, $d = 0$ and $d = 1$ with these probabilities. The probabilities that I's recommendation matches the state of the world are $Pr(\omega = 0 \mid m_P = 0, m_I = 0) = \frac{\beta_P\beta_I}{1-\beta_P-\beta_I+2\beta_P\beta_I}$, and $Pr(\omega = 1 \mid m_P = 0, m_I = 1) = \frac{(1-\beta_P)\beta_I}{\beta_P+\beta_I-2\beta_P\beta_I}$. Then, P's expected utility from playing her equilibrium strategy is $EU_P(s_1 = I, s_2 = 1 \mid m_P = 0) = \alpha(1 - \beta_P) + (\beta_P + \beta_I - 2\beta_P\beta_I) + b$. Consider P's expected utilities from deviating to unilateral actions:

$EU_P(s_1 = 0 \mid m_P = 0) = \alpha\beta_P + 1$, and $EU_P(s_1 = 1 \mid m_P = 0) = \alpha(1 - \beta_P) + b$. It is easy to see that deviating to $s_1 = 1$ is not profitable. To prevent a deviation to $s_1 = 0$, the following inequality must hold:

$$EU_P(s_1 = 0 \mid m_P = 0) \leq EU_P(s_1 = I, s_2 = 1 \mid m_P = 0)$$
$$b \geq \alpha(2\beta_P - 1) + (1 - \beta_P - \beta_I + 2\beta_P\beta_I) \quad (9)$$

Now check to see if P has an incentive to deviate after I made its decision. If $d = 0$, then $EU_P(s_2 = 0 \mid d = 0) = \frac{\alpha\beta_P\beta_I}{1-\beta_P-\beta_I+2\beta_P\beta_I} + 1$, and $EU_P(s_2 = 1 \mid d = 0) = \frac{\alpha(1-\beta_P)(1-\beta_I)}{1-\beta_P-\beta_I+2\beta_P\beta_I} + b$. To prevent a deviation, the following inequality must hold:

$$EU_P(s_2 = 1 \mid d = 0) \geq EU_P(s_2 = 0 \mid d = 0)$$
$$b \geq \frac{\alpha(\beta_P + \beta_I - 1)}{1 - \beta_P - \beta_I + 2\beta_P\beta_I} + 1 \quad (10)$$

If $d = 1$, then $EU_P(s_2 = 0 \mid d = 1) = \frac{\alpha(1-\beta_I)\beta_P}{\beta_P+\beta_I-2\beta_P\beta_I} + 1$, and $EU_P(s_2 = 1 \mid d = 1) = \frac{\alpha(1-\beta_P)\beta_I}{\beta_P+\beta_I-2\beta_P\beta_I} + b + 1$. The deviation is not profitable.

(ii) $\gamma = b$ and $m_P = 1$

Given $m_P = 1$, $\mu'_P = Pr(\omega = 0 \mid m_P = 1) = 1 - \beta_P$. If P takes an issue to I, the probabilities that I receives each message are $Pr(m_I = 0 \mid m_P = 1) = \beta_P + \beta_I - 2\beta_P\beta_I$, and $Pr(m_I = 1 \mid m_P = 1) = 1 - \beta_P - \beta_I + 2\beta_P\beta_I$. Correspondingly, $d = 0$ and $d = 1$ with these probabilities. The probabilities that I's recommendations will match the state of the world are $Pr(\omega = 0 \mid m_P = 1, m_I = 0) = \frac{(1-\beta_P)\beta_I}{\beta_P+\beta_I-2\beta_P\beta_I}$, and $Pr(\omega = 1 \mid m_P = 1, m_I = 1) = \frac{\beta_P\beta_I}{1-\beta_P-\beta_I+2\beta_P\beta_I}$. Then, P's expected utility from playing her equilibrium strategy is $EU_P(s_1 = I, s_2 = 1 \mid m_P = 1) = \alpha\beta_P + (1 - \beta_P - \beta_I + 2\beta_P\beta_I) + b$. Consider P's expected utilities from deviating to unilateral actions: $EU_P(s_1 = 0 \mid m_P = 1) = \alpha(1 - \beta_P) + 1$, and $EU_P(s_1 = 1 \mid m_P = 1) = \alpha\beta_P + b$. Deviating to $s_1 = 1$ is not profitable. To prevent a deviation to $s_1 = 0$, the following inequality must hold:

$$EU_P(s_1 = 0 \mid m_P = 1) \leq EU_P(s_1 = I, s_2 = 1 \mid m_P = 1)$$
$$b \geq \alpha(1 - 2\beta_P) + (\beta_P + \beta_I - 2\beta_P\beta_I) \quad (11)$$

Now check to see if P has an incentive to deviate after I's decision is made. If $d = 0$, then $EU_P(s_2 = 0 \mid d = 0) = \frac{\alpha(1-\beta_P)\beta_I}{\beta_P+\beta_I-2\beta_P\beta_I} + 1$, and $EU_P(s_2 = 1 \mid d = 0) = \frac{\alpha(1-\beta_I)\beta_P}{\beta_P+\beta_I-2\beta_P\beta_I} + b$. To prevent a deviation, the following inequality must hold:

$$EU_P(s_2 = 0 \mid d = 0) \leq EU_P(s_2 = 1 \mid d = 0)$$
$$b \geq 1 + \frac{\alpha(\beta_I - \beta_P)}{\beta_P + \beta_I - 2\beta_P\beta_I} \quad (12)$$

If $d = 1$, then $EU_P(s_2 = 0 \mid d = 1) = \frac{\alpha(1-\beta_P)(1-\beta_I)}{1-\beta_P-\beta_I+2\beta_P\beta_I} + 1$, and $EU_P(s_2 = 1 \mid d = 1) = \frac{\alpha\beta_P\beta_I}{1-\beta_P-\beta_I+2\beta_P\beta_I} + 1 + b$. The deviation is not profitable.

Thus, if inequalities (9)–(12) all hold, we have a partial pooling institutional equilibrium in which the leader defies the institution whenever it recommends policy $x = 0$. It can be shown that condition (10) is the strictest of all. So, if $b \geq 1 + \frac{\alpha(\beta_P+\beta_I-1)}{1-\beta_P-\beta_I+2\beta_P\beta_I}$, then the equilibrium exists; in addition, if and only if $b > 1 + \frac{\alpha(\beta_P+\beta_I-1)}{1-\beta_P-\beta_I+2\beta_P\beta_I}$, the equilibrium is unique.

This completes the proof of Proposition 3.

Proposition 4. *If the voter's belief about a leader's type does not decrease when the voter observes an unexpected consultation of the institution by the leader, then an institutional equilibrium is the only equilibrium to the game.*

Proof. Suppose there is a unilateral equilibrium in which both types only take unilateral actions. Suppose seeing a deviation from the equilibrium path in which a leader takes an issue to the institution, the voter either retains the same belief or increases his belief that the leader is an unbiased type. Then, the deviation will give an unbiased leader at least the same chance of getting reelected. Additionally, the leader's expected utility from policy outcome by taking an issue to the institution is $\alpha\beta_I$, which is greater than the best she could get from a unilateral action ($\alpha\beta_P$). Therefore, the deviation is profitable for an unbiased leader. A contradiction. □

References

Canes-Wrone, Brandice, Michael C. Herron, and Kenneth W. Shotts. 2001. "Leadership and Pandering: A Theory of Executive Policymaking." *American Journal of Political Science* 45(3): 532–50.

Chapman, Terrence L. 2007. "International Security Institutions, Domestic Politics, and Institutional Legitimacy." *Journal of Conflict Resolution* 51(1): 134–66.

Chapman, Terrence L., and Dan Reiter. 2004. "The United Nations Security Council and the Rally 'Round the Flag Effect." *Journal of Conflict Resolution* 48(6): 886–909.

Coate, Stephen, and Stephen Morris. 1995. "On the Form of Transfers to Special Interests." *Journal of Poltical Economy* 103(6): 1210–35.

Dai, Xinyuan. 2002. "Information Systems in Treaty Regimes." *World Politics* 54 (July): 405–36.

Dai, Xinyuan. 2005. "Why Comply? The Domestic Constituency Mechanism." *International Organization* 59 (Spring): 363–98.

Eichenberg, Richard C. 2005. "Victory Has Many Friends." *International Security* 30(1): 140–77.

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

Grieco, Joseph M. 1988. "Anarchy and the Limits of Cooperation: A Realist Critique of the Newest Liberal Institutionalism." *International Organization* 42(3): 485–507.

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.

Haftendorm, Helga, Robert O. Keohane, and Celeste A. Wallander. 1999. *Imperfect Unions: Security Institutions over Time and Space*. New York: Oxford University Press.

Holsti, Ole R. 2004. *Public Opinion and American Foreign Policy*. Ann Arbor: University of Michigan Press.

Keohane, Robert O. 1984. *After Hegemony*. Princeton, NJ: Princeton University Press.

Kull, Steven. 2002a. *Americans on the Conflict with Iraq*. http://www.pipa.org/OnlineReports/Iraq/IraqConflict_Oct02/IraqConflict%20Oct02%20rpt.pdf. PIPA/Knowledge Networks. (accessed May 4, 2006).

Kull, Steven. 2002b. *Americans on Iraq: After the UN Resolution*. http://www.pipa.org/OnlineReports/Iraq/IraqUNRes_Dec02/IraqUNRes%20Dec02%20rpt.pdf. PIPA/Knowledge Networks. (accessed May 4, 2006).

Kull, Steven. 2004. *Americans and Iraq on the Eve of the Presidential Election*. http://www.pipa.org/OnlineReports/Iraq/IraqPresElect_Oct04/IraqPresElect_Oct04_rpt.pdf. PIPA/Knowledge Networks. (accessed May 4, 2007).

Kull, Steven. 2006. *Americans on Iraq: Three Years On*. http://www.worldpublicopinion.org/pipa/pdf/mar06/USIraq_Mar06_rpt.pdf WorldPublicOpinion.org/Knowledge Networks. (accessed May 4, 2007).

Kull, Steven, and I. M. Destler. 1999. *Misreading the Public: The Myth of a New Isolationism*. Washington, DC: Brookings Institution Press.

Mansfield, Edward D., Helen V. Milner, and B. Peter Rosendorff. 2002. "Why Democracies Cooperate More: Electoral Control and International Trade Agreements." *International Organization* 56(3): 477–513.

Martin, Lisa L., and Beth A. Simmons. 1998. "Theories and Empirical Studies of International Institutions." *International Organizations* 52(3): 729–57.

Mearsheimer, John J. 1994–1995. "The False Promise of International Institutions." *International Security* 19(3): 5–49.

Milner, Helen V., and B. Peter Rosendorff. 1997. "Democratic Politics and International Trade Negotiations: Elections and Divided Government as Constraints on Trade Liberalization." *Journal of Conflict Resolution* 41(1): 117–46.

Mitchell, Ronald B. 1998. "Sources of Transparency: Information Systems in International Regimes." *International Studies Quarterly* 42(1): 109–30.

Mitchell, Sara M., and Paul R. Hensel. 2007. "International Institutions and Compliance with Agreements." *American Journal of Political Science* 51(4): 721–37.

Mo, Jongryn. 1995. "Domestic Institutions and International Bargaining: The Role of Agent Veto in Two-Level Games." *American Political Science Review* 89(4): 914–24.

O'Connell, Mary Ellen. 2000. "The UN, NATO, and International Law after Kosovo." *Human Rights Quarterly* 22(1): 57–89.

Oneal, John R., and Bruce Russett. 2003. "Causes of Peace: Democracy, Interdependence, and International Organizations, 1985-1992." *International Studies Quarterly* 47: 371–93.

PIPA/Knowledge Networks. 2003a. *Americans on Iraq & the UN Inspections*. http://www.pipa.org/OnlineReports/Iraq/IraqUNInsp1_Jan03/IraqUNInsp1%20Jan03% 20rpt.pdf. PIPA/Knowledge Networks. (accessed May 4, 2007).

PIPA/Knowledge Networks. 2003b. *Americans on the Iraq War and the Future of the United Nations*. http://www.pipa.org/OnlineReports/Iraq/IraqUNfut_Mar03/IraqUNfut%20Mar03%20rpt.pdf. PIPA/Knowledge Networks. (accessed May 4, 2007).

Putnam, Robert D. 1988. "Diplomacy and Domestic Politics: The Logic of Two-Level Games." *International Organization* 42(3): 427–60.

Ramsay, Kristopher W. 2004. "Politics at the Water's Edge: Crisis Bargaining and Electoral Competition." *Journal of Conflict Resolution* 48(4): 459–86.

Russett, Bruce, John R. Oneal, and David R. Davis. 1998. "The Third Leg of the Kantian Tripod for Peace: International Organizations and Militarized Disputes, 1950-85." *International Organization* 52(3): 441–67.

Schultz, Kenneth. 1998. "Domestic Opposition and Signaling in International Crisis." *American Political Science Review* 94(4): 829–44.

Simmons, Beth. 2000. "International Law and State Behavior: Commitment and Compliance in International Monetary Affairs." *American Political Science Review* 94(4): 819–35.

Smith, Alastair. 1998. "International Crises and Domestic Politics." *American Political Science Review* 92(3): 623–38.

Sobel, Richard. 2000. "To Intervene or Not to Intervene in Bosnia." In *Decisionmaking in a Glass House*, ed. Brigitte L. Nacos, Robert Y. Shapiro, and Pierangelo Isernia. New York: Rowman & Littlefield, 111–31.

Sobel, Richard. 2001. *The Impact of Public Opinion on U.S. Foreign Policy Since Vietnam*. New York: Oxford University Press.

Tago, Atsushi. 2005. "Determinants of Multilateralism in US Use of Force: States of Economy, Election Cycle, and Divided Government." *Journal of Peace Research* 42(5): 585–604.

Tarar, Ahmer. 2001. "International Bargaining with Two-Sided Domestic Constraints." *Journal of Conflict Resolution* 45(3): 320–40.

Voeten, Eric. 2005. "The Political Origins of the UN Security Council's Ability to Legitimize the Use of Force." *International Organization* 59(3): 527–57.

Woodward, Bob. 2004. *Plan of Attack*. New York: Simon & Schuster.