

Should Audiences Cost?

Optimal Domestic Constraints in International Crises*

Scott Ashworth
Harris School of Public Policy Studies
University of Chicago

Kristopher W. Ramsay
Department of Politics
Princeton University

October 4, 2011

Abstract

We study a model in which a leader's crisis behavior and the citizens' responses are both determined optimally. The citizens design incentives with two problems in mind. First, our model identifies a heretofore unnoticed time-consistency problem on the part of leaders who might initiate international crises. Second, we allow the leader and the citizen to disagree about when war is an appropriate policy option. In the optimal scheme, citizens always punish leaders who initiate crises and then back down. Whether they punish leaders for backing down rather than going to war, on the other hand, depends on the status quo and on the costs of war. We identify the conditions under which citizens use re-selection incentives to offset the different preferences over fighting (which involves punishing fighting), and the conditions under which they use "audience costs" to exaggerate that preference divergence (which involves rewarding fighting).

*We thank Serra Boranbay, Ethan Bueno de Mesquita, Insong Kim, Justin Fox, Joanne Gowa, Tom Romer and the seminar audiences at Purdue, Wisconsin, the APSA annual meetings, and the Princeton University Conference on Game Theoretic Analysis of Conflict for helpful comments and suggestions.

Outcomes of international crises can have severe domestic consequences for leaders. Losing can lead to sanctions ranging from removal from office to death. In light of this, it has become a commonplace of international relations scholarship that leaders make decisions with an eye towards their domestic constituencies (Fearon, 1994*a*; Snyder, 1991; Bueno de Mesquita et al., 1992; Schultz, 2001*a*; Goemans, 2000). This literature has yielded many insights into international politics, insights that are not available in perspectives that ignore domestic audiences. But almost all of this literature focuses entirely on the leader’s response to an exogenously fixed pattern of domestic responses, leaving open the question of how endogenously generated incentives affect conflict.

We study a model in which a leader’s crisis behavior and the citizens’ responses are both determined optimally.¹ The leader is both a citizen, directly interested in the outcome of the conflict, and a politician, motivated by retaining the benefits of office. Citizens design incentives for this leader with two problems in mind. First, a leader who might initiate a crisis faces a time-consistency problem: she would like to commit to initiate a crisis only when her state is strong, as this will lead the other state to be generous with concessions. But, with privately known types, generous concessions create a temptation for weak types to enter. Second, we allow the leader and the citizen to disagree about when war is an appropriate policy option—what Jackson and Morelli (2007) call “political bias”. We identify the conditions under which citizens use re-selection incentives to offset this political bias, and the conditions under which they use “audience costs” to exaggerate that political bias.

Our model builds on the standard crisis model, of the sort found in Schultz (1999), Fearon (1994*b*), and Bueno de Mesquita and Lalman (1992). Two countries, Home and Foreign, have a status quo division of some resource. Home’s leader is privately informed about the payoffs from war, and, on the basis of this information, can decide whether or not to challenge the status quo. If challenged, Foreign makes an appeasement offer that

¹Downs and Rocke (1994) also look at the agency problem associated with the decision to go to war and are interested in optimal re-selection rules from the principal’s [voters’] perspective. Their model, however, does not allow for an audience cost interpretation, and they do not fully characterize the optimum for their model.

Home can accept, leading to a new allocation, or reject, leading to a war (modeled as a costly lottery over the resource). The Home leader cares about the outcome of this interaction, and also about being reelected as leader. Reelection is determined by the citizens, creating a political agency model of foreign policy decisions in the canonical form of Barro (1973), Ferejohn (1986), and Austen-Smith and Banks (1989). The model thus combines two canonical models to better understand the links between domestic and international politics.

Our main result is a complete characterization of the optimal incentive scheme for the citizen to implement. There are essentially two classes of optimal incentive scheme. In each, re-election probabilities after backing down are less than those for leaders who maintain the status quo. But audience costs, in the sense of punishing backing down conditional on a crisis, are only sometimes optimal. When the aggregate costs of war are high, the optimal scheme leads to peace for sure. The intuition that citizens should punish leaders for backing because it enhances the home country's bargaining power and leads to better settlements is then correct. But when aggregate costs of war are low, and the optimal scheme features both backing down and war with positive probability, the leader is *rewarded* for backing down rather than for fighting. [Other case here.]

What is the intuition for this result? Take the contrast between the status quo and settlement first. When it chooses its offer, Foreign takes into account the expected severity of a war conditional on the types of Home that initiate conflict. The stronger that Foreign believes Home must be before it is willing to initiate, the more Foreign fears fighting, and thus the more generous Foreign will be in appeasement. This creates a kind of self-control problem for Home: it would like to convince Foreign that it will only initiate conflict when it is strong, so offers will be high. But, since strength is private information, if Home successfully convinces Foreign to be generous, it will face ex-post pressures to start conflict even when it is only moderately strong. So the rich set of appeasement offers introduces a distortion into the leader's decision absent electoral incentives, one that can be solved by punishing backing down relative to the status quo.

We also allow for a second distortion, this time in the decision to fight or take the settlement. If the leader has a private cost/benefit ratio for war that differs from the citizens, then that leader will reject too many offers. This distortion can be corrected by *rewarding* backing down relative to fighting. But it's too quick to leap from that observation to the conclusion that the optimal scheme always rewards backing down relative to fighting. The equilibrium offer from Foreign can also be affected, in equilibrium, by that payoff difference, so the optimal scheme must trade off these two effects. In a model that does not have asymmetric information, Jackson and Morelli (2007) show that citizens always have an incentive to create more biased leaders. In our informationally richer environment, by contrast, the optimum might involve either increasing or decreasing the bias.

These results help sort out the foundations of the literature on audience costs.² Fearon, in his seminal paper, gave an informal optimality-based defense of the assumption:

The [audience cost] results here suggest that . . . , if the principal [read voter] could design a 'wage contract' for the foreign policy agent, the principal would want to commit to punishing the agent for escalating a crisis and then backing down. (Fearon, 1994*a*, p. 581)

While plausible, this defense has not won uniform assent. Schultz, for example, wonders

[w]hy voters would punish their leaders for getting caught in a bluff, if bluffing is sometimes an optimal strategy. After all, anyone who has ever played poker understands that bluffing is not always undesirable behavior. . . . Clearly, additional work remains to be done on why and under what conditions rational voters would impose audience costs. (Schultz, 1999, p. 237, ft. 11)

Our model, in which the voter does optimally design an incentive scheme, shows both conditions under which Fearon's intuition is right, and the limits of that intuition.³

²This work includes theory (Fearon, 1994*a*; Schultz, 1999; Smith, 1998), empirics (Tomz, 2007; Schultz, 2001*b*), and applications to everything from causes of war (Bueno de Mesquita and Siverson, 1995) to dispute settlement in GATT (Busch, 2000) to the role of regional organizations in the consolidation on new democracy (Pevehouse, 2002).

³Smith (1998) and Slantchev (2006) are also interested in equilibrium models of audience costs, but they

We proceed as follows. The next section describes the model, and Section 2 discusses the solution concept. Section 3 characterizes equilibrium for arbitrary re-selection mechanisms. Section 4 discusses the optimal incentive scheme, and Section 5 discusses the robustness of that scheme to our modeling assumptions. The final section concludes.

1 The Model

Consider a world with two countries, *Home* and *Foreign*. We treat *Foreign* as a unitary actor, but split *Home* into a leader, who makes international decisions, and a citizen/voter who decides whether or not to retain the leader.

The countries share a perfectly divisible unit of resource, with *Home*'s status quo share y and *Foreign*'s status quo share $1 - y$. At the outset *Home* has the option to keep the status quo or to demand more, initiating a crisis. If it initiates, *Foreign* gets to propose a new allocation, $(x, 1 - x)$, with $0 \leq x \leq 1$. *Home* can accept this, or reject it and have a war. The war costs *Home*'s citizen c_H and *Foreign* c_F , and the winner takes all of the territory. In a war, the winner is determined by the relative strengths of the countries, a quantity neither knows for sure. Without loss of generality, we can model this with two payoff-relevant states: $\theta \in \{H, F\}$. In state H , *Home* wins a war if it happens, and in state F , *Foreign* wins a war if it happens.

The common prior belief is that the two states of the world are equally likely. Prior to deciding between keeping the status quo and challenging, *Home* observes a signal $s \in [0, 1]$. The signals have conditional densities $f(s \mid \theta = H) = 2s$ and $f(s \mid \theta = F) = 2(1 - s)$. A simple application of Bayes's rule gives the posterior probability that the state is *Home* as $\Pr(\theta = H \mid s) = s$. The prior distribution of this posterior probability is uniform on $[0, 1]$.

consider how rational voters might respond to various crisis strategies of leaders to attempt to screen out decision-makers based on their leadership quality. Those models, therefore, ask if actions that are based on optimal leader selection might be observationally equivalent to audience costs. A similar story comes out of Hess and Orphanides (2001), where leaders also have quality types. We, on the other hand, focus on isolating the agency problem and have only one type of leader quality. This allows us to address the direct question regarding how we might optimally incentivize a leader to use her private information about the outcome from war in the best way, from the citizen's perspective, even if the leader and the citizen have different preferences when it comes to war.

After the crisis, the Home citizen decides to retain or dismiss the Home leader. This decision can be a function of the leader’s observed decisions, but not on the particular value of any settlement or on the outcome of the war.⁴

All players evaluate outcomes based on the final allocation of the territory and whether or not there is a war; in addition, the Home leader prefers retaining office to losing it. To specify payoffs formally, we use the following notation: π is Home’s final share of the territory, w is an indicator function taking the value 1 if there is a war and zero otherwise, and ρ is an indicator function taking the value 1 if the leader is retained and 0 otherwise. Foreign ranks outcomes according to the expectation of $(1-\pi)-wc_F$; the Home citizen ranks outcomes according to the expectation of $\pi - wc_H$; and the Home leader ranks outcomes according to the expectation of $\pi - w\gamma c_H + \rho$, where γ is a parameter less than or equal to 1.

Before turning to the analysis, two of our assumptions require further comment. First, the parameter γ in the Home leader’s payoff measures the degree of conflict of interest between the leader and the citizen (what Jackson and Morelli (2007) call “political bias”). Such conflicts make the leader more eager than the citizen to initiate a war. This is natural when the leader has full access to the spoils of war but does not do the actual fighting (Goemans, 2000; Chiozza and Goemans, 2004; Bueno de Mesquita, Siverson and Woller, 1992). Variation in γ may also reflect institutional differences—compulsory, universal military service, for example, should induce high values of γ (at least for leaders with combat-aged children), while an all volunteer military might insulate leaders more from the costs of war. Either way, this divergence in preferences of citizens and leaders is an important component of our model.

Second, our assumption that reelection can be based on whether or not a settlement occurs but not on its precise details reflects our belief that many real-world settlements will be much too complex for voters to condition on the full details. For example, voters will have a hard time evaluating the consequences of detailed trade negotiations. And it would

⁴In the Section 5, we consider the results if we relax either of these restrictions.

be even harder to evaluate classified concessions of the sort involved in the Cuban missile crisis. In the main presentation of the model, we simplify by assuming the starkest version of this informational asymmetry; Section 5 considers the possibility of partially-settlement-dependent mechanisms.

2 Solution Concept

As this is an extensive game with incomplete information, it is natural to focus on perfect Bayesian-Nash equilibria. But in our context, this is not a very restrictive notion. To see why, consider the voter's choice to reelect or not the incumbent. This choice happens at the end of the game, and does not affect the voter's payoff. Thus any reelection strategy is sequentially rational. This allows us to construct an enormous number of perfect Bayesian equilibria. Technically, we can fix an arbitrary reelection strategy and then solve for an equilibrium of the induced game between the Home leader and Foreign, knowing that the overall profile we end up with will be an equilibrium of the overall game.

This problem is a familiar one in the political agency literature descended from Ferejohn's seminal 1986 contribution. In such a world, leaders have no persistent quality or type and the citizens' payoffs are independent of whether or not the leader is reelected. This means that the voter is indifferent between her two options at every history where she is called upon to keep or dismiss the leader. In fact, Ferejohn shows that the multiplicity is actually a feature of an infinite-horizon agency model, so long as the candidates are all identical. That is, even if the winner of the election would go on to face crises in the future, sequential rationality would not limit equilibrium reelection strategies.

Following the tradition initiated by Ferejohn (1986), we characterize the equilibrium that maximizes the citizen's ex-ante welfare. Think of this as using the following procedure:

1. For each vector of reelection probabilities $\mathbf{r} = (r^{SQ}, r^B, r^W)$ (where r^c is the reelection probability in case c):
 - (a) Solve for (perfect Bayesian) equilibrium behavior as a function of \mathbf{r} .

- (b) Calculate Home's ex-ante payoff in equilibrium, $V(\mathbf{r})$.
2. Choose \mathbf{r} to maximize the function $V(\mathbf{r})$.

We restrict attention to situations in which the optimal scheme is interior: $0 < r^c < 1$ for each c . This corresponds to the more primitive assumption that costs of war are not too large.

We focus on equilibria of a particularly simple form. Say that an entry strategy σ_1 is **monotone** if: Home initiates with signal s' and $s > s'$ together imply that Home initiates with signal s , and say that an equilibrium is a **monotone equilibrium** if the entry decision is monotone. While there might be equilibria that involve non-monotone entry decisions, these profiles would not be equilibria if the model were modified so that entry always carried some very small risk of war, independent of the settlement proposed by Foreign. Thus we focus on monotone equilibria for the remainder of the analysis.

3 Equilibrium for a Fixed Reelection Rule

The first step in implementing the plan announced in the previous section is to characterize equilibrium for a fixed reelection rule. So let \mathbf{r} be arbitrary. How will the crisis unfold?

Start at the end of the game. Home will accept the offer $(x, 1 - x)$ exactly when

$$x + r^B \geq s + r^W - \gamma c_H.$$

Solve this to see that Home accepts x if and only if

$$s \leq \bar{s}(x) \equiv x + (r^B - r^W) + \gamma c_H.$$

Next we consider Foreign's optimal offer. In a monotone equilibrium, Foreign believes that Home's signal, conditional on initiating, is uniform on $[\underline{s}, 1]$ for some \underline{s} . So if Foreign

offers $(x, 1 - x)$, it's payoff is

$$U(x) = \Pr(s \leq \bar{s}(x) \mid s \geq \underline{s})(1 - x) + \Pr(s > \bar{s}(x) \mid s \geq \underline{s})(\mathbb{E}(1 - s \mid s \geq \bar{s}(x)) - c_F). \quad (1)$$

The first term is the settlement times the probability of acceptance, while the second term is the expected payoff conditional on war times the complementary probability.

The function U breaks naturally into three components:

1. If $x \geq 1 - (r^B - r^W) - \gamma c_H$, then the offer is accepted for sure, and $U(x) = 1 - x$.
2. If $x < \underline{s} - (r^B - r^W) - \gamma c_H$, then the offer is rejected for sure, and $U(x) = 1 - \mathbb{E}(s \mid s \geq \underline{s}) - c_F$, a constant.
3. If $\underline{s} - (r^B - r^W) - \gamma c_H < x \leq 1 - (r^B - r^W) - \gamma c_H$, then both acceptance and war have positive probability and U is equal to the quadratic

$$Q(x) = (1 - x) \int_{\underline{s}}^{\bar{s}(x)} \frac{dt}{1 - \underline{s}} + \int_{\bar{s}(x)}^1 (1 - t - c_F) \frac{dt}{1 - \underline{s}}.$$

The basic idea of the optimal offer is then clear: it would be foolish to offer more than $1 - (r^B - r^W) - \gamma c_H$ because that would amount to giving additional territory to an opponent who was going to accept anyway. And given that upper bound, the offer should make Q , whose unconstrained maximum is $\underline{s} + c_F$, as large as possible. This is formalized as:

Lemma 1 *Let*

$$x^* = \min\langle \underline{s} + c_F, 1 - (r^B - r^W) - \gamma c_H \rangle.$$

1. *If $x^* > \underline{s} - (r^B - r^W) - \gamma c_H$, then x^* is the unique optimal offer.*
2. *If $x^* \leq \underline{s} - (r^B - r^W) - \gamma c_H$, then any $x \leq \underline{s} - (r^B - r^W) - \gamma c_H$ is an optimal offer. In particular, x^* is optimal.*

(Proofs omitted from the main text are in Appendix B.)

Given an equilibrium offer (x^*) and private information (s), we can calculate Home's continuation value of starting a crisis as

$$J(s, x^*) = \begin{cases} s + r^W - \gamma c_H & \text{if } s > x^* + (r^B - r^W) + \gamma c_H \\ x^* + r^B & \text{otherwise} \end{cases}.$$

This function is graphed in Figure 1. The dashed lines represent Home's utilities to settlement and war as a function of their private information about success in war. Home will enter if $J(s, x^*) > y + r^{SQ}$, will not enter if $J(s, x^*) < y + r^{SQ}$, and is indifferent if $J(s, x^*) = y + r^{SQ}$. As an immediate implication we have:

Lemma 2 *In any equilibrium in which there is positive probability that Home keeps the status quo and positive probability that Home enters and accepts the offer,*

$$y + r^{SQ} = x^* + r^B.$$

To help show how these components fit together, we compute an equilibrium for the simplest case: $r^{SQ} = r^B = r^W = \gamma = 1$, and $c_H = c_F < y$. H will accept the offer x whenever $s \leq \bar{s}(x) = x + c_H$. Guess that the equilibrium has some (but not all) types initiate and some (but not all) of those accept the appeasement offer. That means we must find a x^* and \underline{s} such that $x^* = \underline{s} + c_F$ and $y = x^*$. That's easy: $x^* = y$ and $\underline{s} = y - c_F$. By the one-shot deviation principle, we now have an equilibrium, *unless* $\underline{s} > \bar{s}(x^*)$, invalidating our guess. But $\bar{s}(x^*) = y + c_H > y - c_F$, so we do in fact have an equilibrium. This situation is illustrated in Figure 2.

Similar constructive arguments can be used in general to establish:

Proposition 1 *Fix any reelection rule \mathbf{r} . The induced crisis game has a pure-strategy monotone equilibrium.*

The proof shows that there are four possibilities for equilibrium behavior:

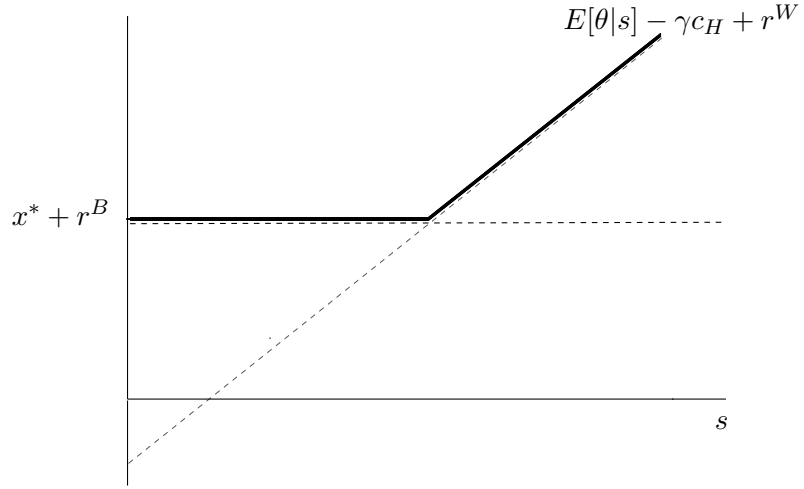


Figure 1: $J(s, x^*)$, in bold, is the upper envelope of the payoff to settlement, $x^* + r^B$, and the payoff to war, $E[\theta|s] - \gamma c_H + r^W$.

1. No types enter;
2. Some types enter and all who do settle;
3. Some types enter and all who do fight; or
4. Some types enter, and some of those settle while others fight.

Next we ask: which of these is optimal?

4 The Optimal Scheme

The optimal scheme is the one that maximizes the citizen's ex-ante payoff. Writing \underline{s} for the least type that initiates, x^* for the offer, and \bar{s} for the least type that fights, this payoff is

$$\underline{s}y + (\bar{s}(x^*) - \underline{s})x^* + \int_{\bar{s}}^1 (t - c_H) dt. \quad (2)$$

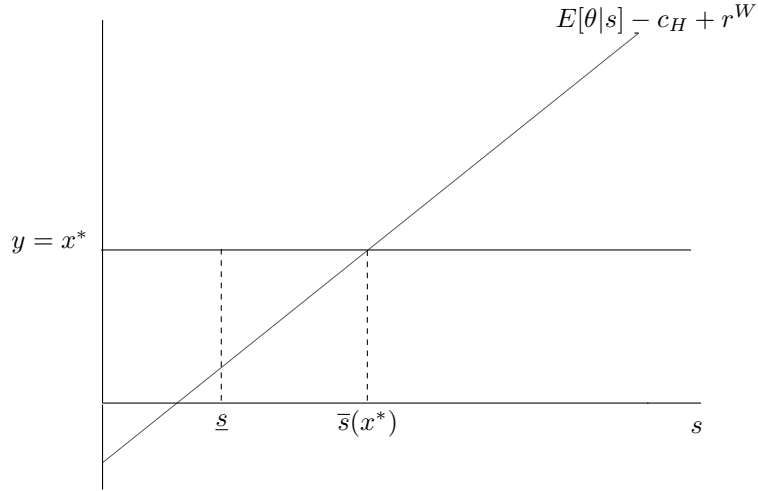


Figure 2: Equilibrium payoffs and cutpoints when there is positive probability of entry and positive probability of accepting the proposed settlement.

The first term is the probability that Home keeps the status quo times Home's status quo share, the second term is the probability that Home's leader initiates a crisis and then accepts the appeasement offer times the value of that offer, and the third term is the expected payoff on the event that there is war. We do not rule out the possibility that $\underline{s} = \bar{s}$ or that one or both of the cutpoints is 0 or 1, so this payoff function covers all of the cases discussed above.

Incentives for fixed x^* As a benchmark, it's helpful to calculate the optimal scheme when x^* is fixed. In this case, the only problem the citizen needs to solve is the discrepancy between her own cost of fighting and the leader's cost of fighting. Comparing this solution to the full solution discussed below will highlight the role of the time-consistency problem in shaping incentives.

Start at the accept/reject decision. The citizen gets $E[\theta|s] - c_H = s - c_H$ from rejection and x^* from acceptance. Thus the appropriate critical type for acceptance is the type that makes the citizen indifferent: $s^* = x^* + c_H$. We can make the leader implement this rule

by setting $r^B - r^W = (1 - \gamma)c_H$. This is quite intuitive—the re-selection differential makes the ruler exactly internalize the extra cost borne by the voter in the case of war.

Now roll back to the entry decision. Write the continuation value for the game conditional on choosing to enter as $\hat{J}(s, x^*) = \max\langle x^*, s - c_H \rangle$. The voter would not enter iff $y \geq \hat{J}(s, x^*)$. Therefore, to implement a play of the game such that the leader of Home does exactly what the citizen would do at each instance, it suffices to take $r^{SQ} - r^B = 0$ and $r^B - r^W = (1 - \gamma)c_H$. Such a scheme looks like this: choose any number $\kappa \in [(1 - \gamma)c_H, 1]$, and set

$$\begin{aligned} r^{SQ} &= \kappa \\ r^B &= \kappa \\ r^W &= \kappa - (1 - \gamma)c_H. \end{aligned}$$

In this benchmark, $r^B > r^W$ —the citizen rewards backing down. This is a direct response to the leader’s political bias. But the leader’s incentives can also be used to affect the other state’s behavior. The optimal scheme in our model responds to exactly that incentive: the citizen manipulates x^* through her choice of \mathbf{r} . Recall that

$$x^* = \min\langle \underline{s} + c_F, 1 - (r^B - r^W) - \gamma c_H \rangle.$$

So there are two ways to manipulate x^* : if some offers are rejected, then the citizen can manipulate x^* only by changing the critical type who enters, while if all offers are accepted, the offer will depend on $r^B - r^W$.

The full optimum To characterize optimal schemes, we maximize the citizen’s payoff (given by (2)), subject to constraints that reflect the dependence on the reelection probabilities, Foreign’s offer, and the Home leader’s war/acceptance decision. It’s then straightforward to write down explicit schemes that lead to the behavior that attains the optimum. As the statement and derivation of the optimal scheme are quite involved, we defer the details

to Appendix A, and focus here on its interpretation. To help elucidate the implications of the optimal scheme, we present the salient implications of the characterization as a series of facts. Each fact follows immediately from the characterization in Proposition 5 in Appendix A.

First we ask what behavior is induced by the optimal scheme. [Replace the fact set up with results or lemmas or something. Find proofs for these claims. They don't seem to be in appendix]

Fact 1 *The optimal scheme always induces positive probability of initiating a crisis and of backing down.*

Thus every optimal scheme generates crises with positive probability. Importantly, the probability that the crisis results in a war varies with the underlying parameters.

Fact 2 *The optimal scheme induces war with positive probability if and only if the total cost of war are low enough ($c_H + c_F < (1/2)(1 - y)$).*

That war is possible only when total costs are low is quite intuitive. War is avoided with probability 1 only when the appeasement offer is so high that all types of Home accept. Since the optimal offer is

$$x^* = \min\langle \underline{s} + c_F, 1 - (r^B - r^W) - \gamma c_H \rangle,$$

we see that increasing F 's cost raises the offer (ignoring the cap), while increasing H 's cost lowers the cap. Both changes tend to make the cap binding.

The possibility of war in the optimal incentive scheme casts doubt on the classical liberal argument that, if the citizens of a country were in control, the country would be peaceful because it is the citizens who pay the “price of war in blood and money” (Kant, 1903; Russett, 1993; Snyder, 1991). That is, even though there is an agency problem between the leader and the citizen, with the leader facing different benefits and costs of starting crises, it is within the power of the citizen to prevent war. What we see is that political control

by the citizens does not eliminate risky behavior, and citizen re-election schemes respond to the risk-reward trade off common in unitary actor models of crisis and war.

Interestingly, the decision about whether or not to use a scheme with positive probability of war is independent of the preference divergence between the leader and the citizen. The form of the optimal scheme, on the other hand, does depend on the preference divergence. This is because the citizen uses the reward for settlement to get the degree of political bias that is optimal for manipulating the offer. How much of that optimal level of political bias needs to come from the reward scheme obviously depends on how much political bias is inherent in the preference divergence.

The optimal degree of political bias depends on what kind of behavior the citizen wants to induce.

Fact 3 *The optimal scheme punishes backing down relative to fighting if and only if the total costs of war are high enough.*

The basic intuition comes from comparing the regions labeled “audience costs” and “internalized costs” in Figure 4. In the audience cost case, the citizen wants to support certain settlement at the highest possible level. To do this, she needs to punish backing down just enough to offset the leader’s cost of fighting. Here the incentives work in a manner reminiscent of the intuitions of Fearon (1994a) and Schultz (2001b): they make backing down costly, stiffen the leader’s stance in bargaining, and lead to a bigger share of the pie. In fact, the optimal audience costs exactly offsets the leader’s cost of fighting, and Home gets everything when it initiates a crisis. (In the space labeled “hybrid regime”, there is no war, but the citizen is unable to design incentives that extract all of the pie.)

In the internalized costs case, on the other hand, the leader will choose to fight with positive probability. And the set of offers he rejects does not affect the offer that is actually made. Thus the citizen wants to offset the existing political bias by rewarding backing down relative to fighting. After the crisis has begun, the citizen wants the leader to internalize fully the cost of war. For $\gamma > 0$, this means making war less attractive than a settlement.

Whether or not backing down is rewarded relative to fighting depends on the underlying

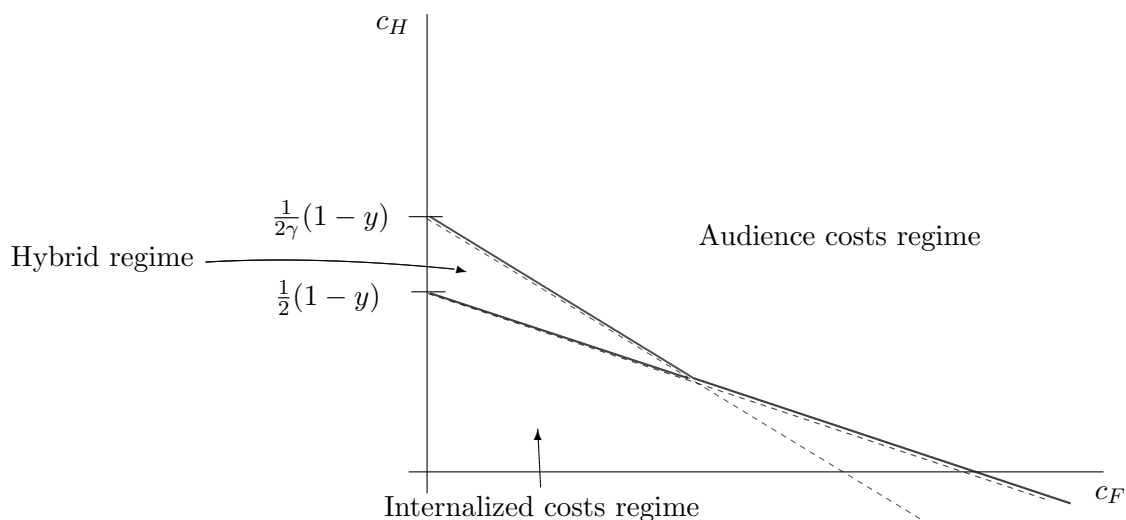


Figure 3: Audience cost regimes as a consequence of citizens' costs of war.

parameters. By contrast, the relative rewards to initiating a crisis and backing down are unambiguously ordered:

Fact 4 *The optimal scheme always punishes backing down relative to the status quo.*

Whatever offer the citizen is trying to extract from F , initiation must be limited. In the internalized cost regime, this is obvious: the offer is an increasing function of the least type who enters. The argument for the audience cost regime is a bit more subtle, but just as intuitive: if F is making a very generous offer, then weak types will want to initiate a crisis to take advantage of it. But if F expects that reaction, it will no longer be willing to make the generous offer. Thus offers can be generous in equilibrium only if accepting those generous offers is costly to leaders. Interestingly, this robust feature of optimal schemes is the one for which Tomz (2007) finds support in his survey experiments on audience costs.

Finally, the model's comparative statics shed some light on when Fearon-Schultz audience costs, i.e. those where backing down leads to a lower probability of reelection than fighting, are a good incentive scheme for citizens to use.

Fact 5 *As Home's status quo share increases, the audience cost regime becomes more likely,*

in the sense that more combinations of c_H and c_F make the audience cost regime optimal.

Proposition 2 *Suppose that there is uncertainty about Home's cost of fighting, then the optimal scheme punishes backing down relative to the status quo and punishes backing down relative to fighting.*

5 Robustness

To keep things simple, we have thought about re-election schemes where selection depends on a coarser set of outcomes than are produced by the extensive form of the game. In particular, we have not allowed the citizens to condition on war outcomes or on the details of the settlement given peace. It is clearly important to consider relaxing these restrictions. We discuss two possible relaxations: schemes where voters can reward winners and losers of wars at different rates and schemes where voters can distinguish between settlements that are better than and worse than the status quo. The optimal incentives characterized above are robust to these extensions.

5.1 Distinguishing winning and losing

Assume that the citizen can distinguish between victory and defeat in the war. The reelection probabilities in these events are denoted r^V and r^D , respectively. Facing such a scheme Home will accept the offer $(x, 1 - x)$ if and only if

$$x + r^B \geq s(1 + r^V - r^D) + r^D - \gamma c_H.$$

Solve this to see that Home accepts x if and only if

$$s \leq \alpha x + \beta$$

where $\alpha = \frac{1}{1+r^V-r^D}$ and $\beta = \frac{r^B-r^D+\gamma c_H}{1+r^V-r^D}$.

The objective of Foreign is then to choose x to maximize

$$\begin{aligned} & (\alpha x + \beta)(1 - x) + \int_{\alpha x + \beta}^1 1 - s - c_F ds & \text{if } \alpha x + \beta \leq 1 \\ & 1 - x & \text{otherwise.} \end{aligned}$$

A simple calculation show the optimal offer is

$$x^* = \min\left(\frac{\underline{s} + \alpha c_F + \beta(\alpha - 1)}{2\alpha - \alpha^2}, \frac{1 - \beta}{\alpha}\right)$$

To see this, note that the second argument of the min function is the point at which the objective switches to $1 - x$, which is strictly decreasing in x .⁵ The first argument is from the first piece of the objective function, the min argument accounts for the utility's piecewise nature.

Now we can give an example to show that *rewarding* losers can increase Home's payoff. Let $c_H = c_F = 1/3$ and let $0 < y < 1/2$. When we constrain $r^V = r^D$, our previous results show that the optimal scheme induces entry but no war, and give Home's citizen a payoff

$$(1 - c_F)y + c_F = (2/3)y + 1/3.$$

To show that this is no longer optimal, we do not have to solve the complete optimization problem; it suffices to display a sample scheme that's better. Since the optimal scheme with unrestricted \mathbf{r} must give payoffs at least this great, this shows that allowing leaders to be rewarded for losing can make the citizen better off.

Here's the scheme. Set $\beta = 0$ and have everyone enter, i.e. set $\underline{s} = 0$. Then the equilibrium offer from Foreign is

$$x^* = \min\left(\frac{c_F}{2 - \alpha}, \frac{1}{\alpha}\right).$$

The first argument of min is strictly increasing in α while the second argument is decreasing

⁵For future reference, note that rewarding losers, $r^V < r^D$, is equivalent to $\alpha > 1$.

in α , so x^* is maximized where they are equal:

$$\frac{1/3}{2-\alpha} = \frac{1}{\alpha}.$$

Solving this equation gives $\alpha = 3/2$. Since every type of Home enters and accepts the offer, Home's payoff is $x^* = 2/3$, which is more than $(2/3)y + 1/3$ since $y < 1/2$. As $\alpha > 1$, this scheme rewards losers.

The idea of rewarding losers seems very strange, but there is actually a compelling intuition for the idea. The leader's payoff to war as a function of his signal is

$$s(1 + r^V) + (1 - s)r^D - \gamma c_H = (1 + r^V - r^D)s + r^D - \gamma c_H.$$

This is, of course, increasing in strength measured by s . The rate of this increase with s is governed by the difference $r^V - r^D$: the smaller is this difference, the less does the war payoff increase with s . This implies the fraction of types bought out of war when Foreign increases the offer by a fixed increment is greater the lower is the difference $r^V - r^D$. Consequently, increasing the reward to losing (relative to the reward to winning) increases Foreign's marginal incentive to make larger offers.

So, within the current model, rewarding losers makes a lot of intuitive sense. But doing so may be a bad idea for reasons neglected by the model. For example, rewarding losers of wars might create perverse incentives for leaders to mismanage conflicts.⁶ Thus we turn to the question of what is the optimal scheme subject the constraint that $r^V \geq r^D$. We show that the answer to that question is identical to the answer we derived before when the citizen could not make rewards contingent on the war's outcome.

Proposition 3 *Assume the citizen can distinguish victory from defeat, but is constrained to reelect victors at least as often as losers ($r^V \geq r^D$). The optimal incentive scheme is that described in Appendix A.*

⁶While it is not obvious that the leader has bad incentives when losers are rewarded—he prefers to win as long as $1 + r^V > r^D$, which is consistent with $r^V < r^D$ —no one has worked out the incentives to manage a war and, therefore, we cannot be sure of the consequences.

5.2 Conditioning on the offer

Next we consider the consequences of letting the reelection probability partially depend on the offer that the leader accepts. To do so, let r^B depend on the offer x . There are a number of ways in which r^B can depend on the settlement x . We first observe that we could find a potentially complex re-selection scheme that varies with the size of the settlement. Such a fine grained scheme requires much of the citizens and assumes they know many details of the agreement. Moreover, as the optimal scheme may not even be continuous, the difficulty of characterizing such a scheme would take us too far afield in this analysis. This does not mean we can say nothing about re-election schemes that depend on the size of the peaceful settlement. Consider the case where the citizen can at least partly condition the retention decision on the value of an accepted settlement. In particular, assume that she can tell whether a settlement represents an improvement on the status quo or not. We claim that this does not affect the optimal scheme.

Proposition 4 *Assume the citizen can distinguish settlements that are better than the status quo from those that are worse than the status quo. The optimal incentive scheme is that described in Appendix A.*

The intuition is simple. The optimal scheme that does not condition on the offer induces an offer that is better than the status quo for Home. Changing the reward to accepted offers worse than the status quo thus has no impact on the local optimality of the offer for Foreign. All that changing those rewards can do is to make significantly lower offers more acceptable to Home, a change which can only reduce the optimal offer for Foreign.

Together Propositions 3 and 4 show that we can relax both the requirement that a scheme not depend on the outcome of a war, as long as winners do at least as well as losers, and the assumption that schemes not depend on the actual value of a settlement and still get the same characterization as Proposition 1. That is, there will essentially be two qualitatively different regimes, one with zero probability of war and reelection probabilities at the status quo and fighting with higher values than backing down (i.e., audience costs) and one with positive probability of war and cost internalization that rewards backing down

relative to fighting. That is, the optimal scheme and the incentives driving it are robust to these significant changes in the modeling assumptions.

6 Conclusion

In this paper we have characterized the optimal incentive scheme for a citizen to give to a ruler who might engage in a simple form of crisis bargaining. Doing so allows us to endogenize the domestic political constraints that have played such a large role in recent IR theory. These political constraints are sensitive to two different agency problems. The first, which is well-explored in the literature, is the divergence between the leader's private payoff to war and the citizens' payoffs to war. The second, which to our knowledge is new, is a kind of time-consistency problem faced by the leader. This consistency problem comes from the fact that committing to keep the status quo unless the private information received by the leaders is quite favorable for Homes prospects in war leads to high appeasement offers in the case of initiation, but those same high offers make it attractive to initiate crisis when their signal is not so favorable.

The citizen's response to the time-consistency problem is quite robust: the leader always faces a lower probability of reelection after backing down than if she keeps the status quo. This allows a citizen to solve the leader's time-consistency problem. But the incentive we thought we knew well, the decision to enhance or offset the leader's political bias is more sensitive to the environment. When total costs of war are low, the citizen offsets the leader's political bias by rewarding backing down more than going to war. This leads the leader to fully internalize the costs of fighting, an action she nonetheless takes with positive probability in the equilibrium. When total costs of fighting are large, on the other hand, the citizen designs a scheme that leads to peace with probability one. In this case, the citizen enhances the bargaining power of the leader by punishing her for backing down rather than fighting.

While the model is much too stylized to capture the full richness of empirical discussions of audience costs, it does help clear up an ambiguity in the literature's treatment of the

issue. In Fearon’s (1994*a*) canonical theoretical discussion, as well as in the cases discussed by Schultz (2001*b*), audience costs refer to punishment for backing down once a crisis has started. Another take on audience costs, one tested in Tomz’s (2007) survey experiments, is that they are costs associated with initiating and then backing down.⁷ In terms of our model, Fearon and Schutz are discussing the contrast between reelection probabilities conditional on backing down and on war, while Tomz is talking about the contrast between reelection probabilities conditional on backing down and on keeping the status quo from the beginning. The model highlights that these are conceptually different parts of the optimal incentive scheme, and that whether or not each is used responds to different aspects of the agency problem.

There is clearly much more work to be done before we have a full picture of leader and citizens’ incentives for the agency problem during a crisis. One obvious question, which we are attempting to answer in ongoing research, is how sensitive are the results to the precise form of uncertainty? Our model has asymmetric information about the relative strength of the two states. Few formal IR models follow such an approach; most feature asymmetric information about resolve (operationalized as uncertainty about costs of fighting). Since Fey and Ramsay (2009) show that the difference can be consequential, it is interesting to think about optimal incentives in a model with known strength but uncertain resolve.

It’s also easy to think of more substantial extensions. The pure moral hazard approach we have adopted allows the citizen to exercise substantial commitment power. As Fearon (1999) forcefully argues, that commitment power will be substantially curtailed if there is heterogeneity in the pool of potential leaders. (Smith (1998) and Slantchev (2006) have constructed adverse selection models of audience costs in which commitments are not possible.) An interesting extension would ask what kinds of political institutions would best help manage these commitment problems and allow the citizen to get close to the optimal scheme we characterize here. We also hope to be able to analyze a more symmetric model, one in which both leaders are given incentives by their citizens. These extensions create a

⁷Fearon (1994*a*) comments on p.585 that audience costs may “affect entry decisions” and discusses how his model must be modified to consider this dimension of audience costs.

rich menu of options for further research.

A The Optimal Scheme

Proposition 5 1. If $c_H \leq \frac{1}{2}(1 - y - c_F)$ then a scheme is optimal if and only if it is of the form

$$\begin{aligned} r^{SQ} &= \kappa \\ r^B &= \kappa - 1 + y \\ r^W &= \kappa - 1 + y + \gamma c_H \end{aligned}$$

for some κ .

2. If $c_H \geq \max(\frac{1}{2\gamma}(1 - y - c_F), \frac{1}{2}(1 - y - c_F))$, then a scheme is optimal if and only if it is of the form

$$\begin{aligned} r^{SQ} &= \kappa \\ r^B &= \kappa - (c_H + c_F) \\ r^W &= \kappa - ((2 - \gamma)c_H + c_F) \end{aligned}$$

for some κ .

3. If $\frac{1}{2}(1 - y - c_F) < \frac{1}{2\gamma}(1 - y - c_F)$ and $\frac{1}{2}(1 - y - c_F) < c_H \leq \frac{1}{2\gamma}(1 - y - c_F)$, then a scheme is optimal if and only if it is of the form

$$\begin{aligned} r^{SQ} &= \kappa \\ r^B &= \kappa - \frac{1}{2}(1 - y) - \frac{1}{2}c_F \\ r^W &= \kappa - (1 - y) - c_F + \gamma c_H \end{aligned}$$

for some κ .

Proof To ease the notational burden, define $\beta = r^B - r^W$. We study the program

$$\begin{aligned}
\max_{\underline{s}, \beta, x} \quad & \underline{s}y + (\bar{s} - \underline{s})x + \int_{\bar{s}}^1 (t - c_H) dt. \\
\text{st} \quad & \bar{s} = \min\langle 1, x + \beta + \gamma c_H \rangle \\
& x \leq \underline{s} + c_F \\
& x \leq 1 - \beta - \gamma c_H \\
& 0 \leq x \leq 1.
\end{aligned}$$

In this program, acceptance decisions are optimal for Home by construction. The difference $r^{SQ} - r^B$ is unrestricted by the solution, so we can choose it to make all types indifferent between the status quo and the offer. This does not impose all of the restrictions implied by equilibrium—we have not imposed, for example, $\underline{s} \leq \bar{s}$. Our approach is to show that solutions to this program are actually supportable as equilibria, which implies that they are in fact solutions to the problem of optimizing over the equilibrium correspondence.⁸

Direct calculation gives:

Lemma 3 *If the second inequality constraint is satisfied, then $\bar{s} = x + \beta + \gamma c_H$.*

Thus we can rewrite the program as

$$\begin{aligned}
\max_{\underline{s}, \beta, x} \quad & \underline{s}y + (x + \beta + \gamma c_H - \underline{s})x + \int_{x + \beta + \gamma c_H}^1 (t - c_H) dt. \\
\text{st} \quad & x \leq \underline{s} + c_F \\
& x \leq 1 - \beta - \gamma c_H \\
& 0 \leq x \leq 1.
\end{aligned}$$

Lemma 4 *At the solution to the program, the first inequality constraint must bind.*

Proof It's obvious that the solution neglecting both inequality constraints is to set $x = 1$,

⁸It might seem that our constraints are too restrictive—Lemma 1 allows offers greater than x^* for some reelection probabilities. But notice that that can only happen when the offer is sure to be rejected, so setting the offer arbitrarily to x^* in such cases does not miss any possible payoffs to the citizen.

$\underline{s} = 0$, and $\beta = -\gamma c_H$. But that violates the first inequality constraint.

Now assume that the second inequality constraint binds. In this case, the offer is $x^* = 1 - \beta - \gamma c_H$. Given that, the equality constraint implies $\bar{s} = 1$, and the payoff reduces to

$$\underline{s}y + (1 - \underline{s})x.$$

If the first inequality constraint does not bind, then the solution must be for $x = 1$ and $\underline{s} = 0$. But then $\underline{s} + c_F = c_F < 1 = x$, contradicting the claim that the first constraint is slack. \square

Now we solve the optimization program ignoring the second inequality constraint. If the solution turns out to satisfy that neglected constraint, it will actually be the solution to the full program. Since the first inequality constraint must bind, we can substitute to get the program

$$\begin{aligned} \max_{\beta, x} \quad & (x - c_F)y + (\beta + \gamma c_H + c_F)x + \int_{x+\beta+\gamma c_H}^1 (t - c_H) dt \\ \text{st } & 0 \leq x \leq 1. \end{aligned}$$

The first-order condition for β simplifies to

$$\beta = (1 - \gamma)c_H.$$

Ignoring the boundary constraints on x , the maximizing value of x is $x = y + c_H + c_F$. Thus the optimal x is

$$x = \min\langle 1, y + c_H + c_F \rangle.$$

To implement this, recall from Lemma 3, the leader must be indifferent between the status quo and the offer: $y + r^{SQ} = x + r^B$, or

$$r^{SQ} - r^B = x - y > 0.$$

Therefore, there are costs for settlement relative to the status quo, but not costs relative to going to war.

Finally, we check the neglected inequality constraint: $x \leq 1 - \beta - \gamma c_H$. Notice first that this can never be satisfied if $x = 1$, so we only need to consider the interior solution. Substituting the values that solve the relaxed program gives $y + c_H + c_F \leq 1 - c_H$, or

$$c_H \leq \frac{1}{2}(1 - y - c_F). \quad (3)$$

That gives us case 1 in the Proposition.

When inequality (3) is not satisfied, both inequality constraints must bind. Substituting them into the reduced objective gives the unconstrained program

$$\begin{aligned} \max_x \quad & (x - c_F)y + (1 - x + c_F)x \\ \text{st } & 0 \leq x \leq 1. \end{aligned}$$

Neglecting the boundary constraints, the first-order condition for this program is

$$y + 1 - 2x + c_F = 0.$$

Thus the actual solution to the program is

$$x = \min \left\langle 1, \frac{1}{2}(1 + y + c_F) \right\rangle.$$

This gives us two subcases. If $y + c_F \geq 1$, then the optimal scheme has

$$1 = 1 - \beta - \gamma c_H,$$

or

$$\beta = -\gamma c_H < 0;$$

and $y + r^{SQ} = 1 + r^B$, or $r^{SQ} - r^B = 1 - y$. This is case 2 of the Proposition.

The other subcase is when $y + c_F < 1$. Then we back out β from

$$\frac{1}{2}(1 + y + c_F) = 1 - \beta - \gamma c_H,$$

which gives

$$\beta = \frac{1}{2}(1 - y - c_F) - \gamma c_H.$$

Similarly, we use $y + r^{SQ} = x + r^B$ to get

$$r^{SQ} - r^B = \frac{1}{2}(1 - y) + \frac{1}{2}c_F.$$

This is case 3 of the Proposition. □

B Omitted Proofs

B.1 Equilibrium for Fixed Incentives

Proof of Lemma 1 Evaluating the integral shows that Q is quadratic with the coefficient on x^2 is $-\frac{1}{2}$, so Q is concave. Since the other two components of Q are linear, this shows that U is strictly concave on $(\underline{s} - (r^B - r^W) - \gamma c_H, \infty)$ and constant otherwise, so it is globally quasiconcave. Furthermore, U is continuous and is differentiable except possibly at $x = \underline{s} - (r^B - r^W) - \gamma c_H$.

The next step is to look more closely at Q as a function on all of \mathbb{R} . Multiply through by $(1 - \underline{s})$ and differentiate (remembering that $\bar{s}'(x) = 1$) to get

$$(1 - \underline{s})Q'(x) = -(\bar{s}(x) - \underline{s}) + (1 - x) - (1 - \bar{s}(x) - c_F).$$

Equate this to 0 and solve to see that Q is maximized at $x = \underline{s} + c_F$.

This means that U is nondecreasing up to

$$x^* = \min(\underline{s} + c_F, 1 - (r^B - r^W) - \gamma c_H),$$

and is strictly decreasing thereafter. Thus x^* is always an optimal offer. If $x^* > \underline{s} - (r^B - r^W) - \gamma c_H$, then x^* is the unique optimizer. If $x^* \leq \underline{s} - (r^B - r^W) - \gamma c_H$, on the other hand, then any $x \leq \underline{s} - (r^B - r^W) - \gamma c_H$ is optimal. \square

Proof of Proposition 1 We proceed by cases, constructing an equilibrium in each.

1. $y + r^{SQ} \geq 1 - \gamma c_H + r^W$. In this case, there is an equilibrium in which no type enters. To support this, assume that F responds to (off-path) entry with beliefs that the entrant is type $s = 1$ with probability 1, and so must offer $x^* = 1 - (r^B - r^W) - \gamma c_H$. By construction, this offer makes the type $s = 1$ type indifferent between acceptance and war. And since war is worse than the status quo, that means entry is worse than the status quo. All lower types have continuation values bounded above by the highest types, so no type enters.
2. $y + r^{SQ} \leq -\gamma c_H + r^W$. In this case, there is an equilibrium in which every type enters. If everyone enters, the offer is $x^* = \min\langle c_F, 1 - (r^B - r^W) - \gamma c_H \rangle$. This creates two subcases. If $c_F \geq 1 - (r^B - r^W) - \gamma c_H$, then every type enters and accepts the offer. If the inequality is reversed, then every type enters, and types $s < c_F + (r^B - r^W) + \gamma c_H$ accept.
3. $-\gamma c_H + r^W < y + r^{SQ} < 1 - \gamma c_H + r^W$. Again we must consider subcases.
 - (a) $y + r^{SQ} \leq \min\langle c_F, 1 - (r^B - r^W) - \gamma c_H \rangle + r^B$. There is an equilibrium in which every type enters, just as in case 2.
 - (b) $y + r^{SQ} > \min\langle c_F, 1 - (r^B - r^W) - \gamma c_H \rangle + r^B$ and $(r^W - r^B) - \gamma c_H > c_F$. Let \hat{s} be the type that is indifferent between the status quo and war:

$$\hat{s} = y + r^{SQ} + \gamma c_H - r^W.$$

We will find an equilibrium in which type s enters exactly when $s \geq \hat{s}$, and all types that enter reject the offer in favor of fighting. By construction, no entering

type wants to deviate to the status quo and no type that does not enter wants to deviate and fight. To complete the equilibrium we must characterize the offer and show that no type wants to deviate to accept it. But that's easy—the offer is $x^* = \hat{s} + c_F$. So we need

$$y + r^{SQ} = \hat{s} - \gamma c_H + r^W \geq \hat{s} + c_F + r^B.$$

But this is equivalent to our defining condition, $(r^W - r^B) - \gamma c_H > c_F$.

(c) $y + r^{SQ} > \min\langle c_F, 1 - (r^B - r^W) - \gamma c_H \rangle + r^B$ and $(r^W - r^B) - \gamma c_H \leq c_F$. Let \hat{s} be the type that is indifferent between the status quo and war:

$$\hat{s} = y + r^{SQ} + \gamma c_H - r^W.$$

We will find an equilibrium in which type s fights exactly when $s \geq \hat{s}$. First, try for an equilibrium in which some types take the appeasement offer while others fight. In such an equilibrium, we must have $y + r^{SQ} = x^* + r^B$, by Lemma 2. Since some types are fighting, we have $x^* = \underline{s} + c_F$, so we can substitute and solve to get

$$\underline{s} = y + r^{SQ} - c_F - r^B.$$

Since accepting the offer and the status quo have the same payoff, the offer is accepted by types less than \hat{s} . The observations give an equilibrium as long as $\underline{s} \leq \hat{s}$, or

$$y + r^{SQ} - c_F - r^B \leq y + r^{SQ} + \gamma c_H - r^W.$$

This simplifies to

$$(r^W - r^B) - \gamma c_H \leq c_F,$$

which is true.

□

B.2 Robustness

Proof of Proposition 3 It is again legitimate to focus on the two cases of entry but no war and entry with some but not all types settling. No entry was shown already to be dominated by a scheme with $r^V = r^D$, and that dominating scheme is still feasible here. Similarly, a scheme with entry but no settlement implies that the weakest type who enters is passing up an offer that dominates (from the citizen's point of view) war for that type, something the optimal scheme will rule out.

Turn now to the case of entry but no war. Writing $x^*(\underline{s}, \alpha, \beta)$ for the optimal offer given \underline{s} , α , and β , the citizen's payoff is

$$\underline{s}y + (1 - \underline{s})x^*(\underline{s}, \alpha, \beta).$$

At an unconstrained optimum, the first order conditions include

$$\begin{aligned} 0 &= (1 - \underline{s}) \frac{\partial x^*}{\partial \beta} \\ &= (1 - \underline{s}) \frac{\alpha - 1}{2\alpha - \alpha^2}. \end{aligned}$$

Since some entry means $1 - \underline{s} \neq 0$, this FOC implies $\alpha = 1$, or $r^V = r^D$.

Finally, notice that in ignoring the kink in the in Foreign's objective, we have actually allowed Home deviations that give higher payoffs than those she can actually achieve. The fact that Home does not take any of those is more than enough to show that it will not take any of the deviations that are available.

Now consider the case with entry and both settlement and war. We can write down the citizen's new optimization program, with possibly differing r^V, r^D , and also applying the constraint that $r^V \geq r^D$. Applying the Kuhn-Tucker Theorem, we know that any solution necessarily satisfies a series of first order and complementary slackness conditions.

$$\begin{aligned}
& \frac{\underline{s}^2 - \underline{s}\alpha^2\delta + c_h\alpha^2\underline{s} + \alpha^2c_f^2 + 2c_h\alpha^2c_f - \alpha^2\delta^2 + 2c_h\alpha^2\delta - \alpha\underline{s}^2}{\alpha^2(\alpha-2)^2} - \lambda_1 + \lambda_2 = 0, \\
& \frac{\alpha\delta + \underline{s}\alpha - c_h\alpha - \underline{s}}{\alpha-2} = 0, \\
& \frac{y\alpha^2 + \alpha^2\delta - 2y\alpha - \alpha\delta - c_h\alpha + \underline{s}}{\alpha(\alpha-2)} = 0, \\
& \lambda_1(1-\alpha) = 0, \\
& \lambda_2(\alpha - \frac{1}{2}) = 0,
\end{aligned}$$

where the final two equations are the complementary slackness conditions, and where we require that $\lambda_i \geq 0$, $i = \{1, 2\}$.

To prove the result we first show that at least one of the two constraints bind. To see this suppose not. That is, consider the case where neither constraint binds, meaning $\lambda_1 = \lambda_2 = 0$. We then have

$$\begin{aligned}
& \frac{\alpha\delta + \underline{s}\alpha - c_h\alpha - \underline{s}}{\alpha-2} = 0 \\
& \frac{y\alpha^2 + \alpha^2\delta - 2y\alpha - \alpha\delta - c_h\alpha + \underline{s}}{\alpha(\alpha-2)} = 0.
\end{aligned}$$

The solution to these two equations, given some chosen α , is

$$\begin{aligned}
\delta &= \frac{c_h + y(1-\alpha)}{\alpha}, \\
\underline{s} &= y + c_h.
\end{aligned}$$

Substituting these values back into the first order conditions and simplifying leads to

$$\frac{c_h^2 + c_f^2 + 2c_h c_f}{(\alpha-2)^2}.$$

Observe that because war costs are strictly positive this first order condition can never be satisfied, contradicting that there exists a solution to the scheme where neither constraint

binds.

Now we are left with the task of evaluating the program when the constraints bind. This gives two possible solutions to consider. Suppose constraint (2) binds, implying $\alpha = 1/2$. Solving the system with $\alpha = 1/2$, we see the only solution gives $\lambda_2 < 0$ contradicting the Kuhn-Tucker conditions. This leaves the case where $\alpha = 1$ and the solution

$$\begin{aligned}\alpha &= 1, \\ \delta &= c_h, \\ \lambda_1 &= c_h^2 + c_f^2 + 2c_h c_f, \\ \lambda_2 &= 0, \\ \underline{x} &= y + c_h.\end{aligned}$$

But this solution is the solution to the program in Proposition 2, where the leader is rewarded for backing down relative to war, completing the proof. \square

Proof of Proposition 4 Consider starting from a scheme that makes the two backdown rewards equal. The optimal such has $x^* > y$. This means that the function $U(x)$ defined in Equation 1 attains it's maximum at $x^* > y$. What happens if we then change the backdown rewards? Changing the reward for backing down with a settlement worse than the status quo leaves U unchanged for $x > y$, but introduces a jump discontinuity at $x = y$. This can only lower the offer. Similarly, changing the reward for backing down with an advantageous settlement causes either the offer to jump down to the discontinuity or to change in a way that was feasible in the baseline case. If the change was feasible in the baseline, then it does not improve on the baseline case's solution, since that was already optimal. And reducing the offer can't be optimal either: then the relaxed program from the proof of Proposition A would still give an upper bound on the citizen's payoff, and we know that upper bound can be attained without conditioning on the direction of change from the status quo.

References

- Austen-Smith, David and Jeffrey Banks. 1989. Electoral Accountability and Incumbency. In *Models of Strategic Choice in Politics*, ed. Peter C. Ordeshook. University of Michigan Press.
- Barro, Robert J. 1973. "The Control of Politicians: An Economic Model." *Public Choice* 14:19–42.
- Bueno de Mesquita, Bruce and David Lalman. 1992. *War and Reason: Domestic and International Imperative*. New Haven, CT: Yale University Press.
- Bueno de Mesquita, Bruce, James D. Morrow, Randolph M. Siverson and Alastair Smith. 1992. *War and Reason: Domestic and International Imperative*. New Haven, CT: Yale University Press.
- Bueno de Mesquita, Bruce and Randolph M. Siverson. 1995. "War and the Survival of Political Leaders: A Comparative Study of Regime Types and Political Accountability." *American Political Science Review* 89(4):841–855.
- Bueno de Mesquita, Bruce, Randolph M. Siverson and Gary Woller. 1992. "War and the Fate of Regimes: A Comparative Analysis." *American Political Science Review* 86(3):638–646.
- Busch, Marc L. 2000. "Democracy, Consultation, and the Paneling of Disputes under GATT." *The Journal of Conflict Resolution* 44(4):425–446.
- Chiozza, Giacomo and H.E. Goemans. 2004. "International Conflict and the Tenure of Leaders: Is War Still *Ex Post* Inefficient?" *American Journal of Political Science* 48(3):504–619.
- Downs, George W. and David M. Rocke. 1994. "Conflict, Agency, and Gambling for Resurrection: The Principle-Agent Problem Goes to War." *American Journal of Political Science* 38(2):362–380.

- Fearon, James D. 1994a. "Domestic Political Audiences and Escalation of International Disputes." *American political Science Review* 88(3):577–592.
- Fearon, James D. 1994b. "Signaling versus the Balance of Power and Interests: An Empirical Test of a Crisis Bargaining Model." *Journal of Conflict Resolution* 38(2):236–269.
- Fearon, James D. 1999. Electoral Accountability and the Control of Politicians: Selecting Good Types versus Sanctioning Poor Performance. In *Democracy, Accountability, and Representation*, ed. Przeworski, Stokes and Manin. Cambridge University Press.
- Ferejohn, John. 1986. "Incumbent Performance and Electoral Control." *Public Choice* 50(1/3):5–25.
- Fey, Mark and Kristopher W. Ramsay. 2009. "Uncertainty and Incentives in Crisis Bargaining Games." *University of Rochester, Typescript* .
- Goemans, Hein E. 2000. *War and Punishment: The Causes of War termination and the First World War*. Princeton, NJ: Princeton University Press.
- Hess, Gregory D. and Athanasios Orphanides. 2001. "War and Democracy." *The Journal of Political Economy* 109(4):776–810.
- Jackson, Matthew O. and Massimo Morelli. 2007. "Political Bias and War." *American Economic Review* 97(4):1353–1373.
- Kant, Immanuel. 1903. *The Perpetual Peace; A Philosophical Essay*. London, UK: Swan Sonnenschein & Co LIM.
- Pevehouse, Jon C. 2002. "With a Little Help from My Friends? Regional Organizations and the Consolidation of Democracy." *American Journal of Political Science* 46(3):611–626.
- Russett, Bruce. 1993. *Grasping the Democratic Peace: Principles for a Post-Cold War World*. Princeton, NJ: Princeton University Press.

- Schultz, Kenneth. 1999. "Do Democratic Institutions Constrain or Inform? Two Contrasting Institutional Perspectives on Democracy and War." *International Organization* 53(2):233–266.
- Schultz, Kenneth. 2001*a*. *Democracy and Coercive Diplomacy*. New York: Cambridge University Press.
- Schultz, Kenneth. 2001*b*. "Looking for Audience Costs." *Journal of Conflict Resolution* 45(1):32–60.
- Slantchev, Branislav L. 2006. "Politicians, the Media, and Domestic Audience Costs." *International Studies Quarterly* 50(2):445–477.
- Smith, Alastair. 1998. "International Crises and Domestic Politics." *American Political Science Review* 92(3):623–638.
- Snyder, Jack. 1991. *Myths of Empire: Domestic Politics and International Ambition*. Ithica, NY: Cornell University Press.
- Tomz, Michael. 2007. "Domestic Audience Costs in International Relations: An Experimental Approach." *International Organization* 61(4):821–840.