Audience Costs and the Dynamics of War and Peace*

Conditionally accepted at the *American Journal of Political Science*

Casey Crisman-Cox† Michael Gibilisco‡

Abstract

We estimate country-specific audience costs and examine their substantive effects on the evolution of interstate disputes. Unlike past efforts at estimating audience costs, our approach uncovers these estimates without replying on proxy variables (e.g., democracy) by using an infinitely repeated and dynamic game of crisis escalation. Contrary to intuition, our results show that increases in a country's audience costs encourage it to initiate disputes in equilibrium, because the costs serve as a commitment device during the subsequent crisis, incentivizing the country to stand firm and coercing its rival to back down. Nonetheless, the results demonstrate that larger audience costs would result in more peace worldwide, as they also discourage potential opponents from initiating disputes. Beyond regime type, we find that a free press, provisions for executive appointment or removal, and historical rivalries are also important determinants of audience costs.

Word Count: 9,729

*Thanks to Scott Abramson, Rob Carroll, Kevin Clarke, Allan Dafoe, Mark Fey, Tasos Kalandrakis, Brenton Kenkel, Bethany Lacina, Jack Paine, Curt Signorino for helpful comments and suggestions. Earlier versions of the paper also benefited from audiences at the 2015 SPSA meeting, the 2015 PolMeth conference, the 2015 EPSA meeting, the 2016 APSA meeting and the University of Rochester. All remaining errors are the fault of the authors. Replication code and data will be available on GitHub upon publication.

†Washington University in St. Louis. Email: c.crisman-cox@wustl.edu

‡University of Rochester. Email: michael.gibilisco@rochester.edu

1 Introduction

Audience costs—or the costs leaders pay from backing down before their opponents in international disputes—are ubiquitous in interstate conflict theory. Scholars use them to explain a variety of phenomena including crisis bargaining, war duration, economic sanctions, and the democratic peace. This prominence has sparked two fervid debates: Do audience costs exist and how would we know? Specifically, the former question asks to what extent are leaders punished for backing down (e.g., Snyder and Borghard 2011; Tomz 2007), while the latter concerns the appropriate methodology for answering such a question (e.g., Kurizaki and Whang 2015; Partell and Palmer 1999; Schultz 2001).

Two substantial impediments prevent progress in either debate. First, since Fearon's (1994*a*) canonical paper, researchers traditionally proxy audience costs using democracy scales such as polity2, and these measures often determine case-selection in qualitative studies and independent variables in quantitative ones (Partell and Palmer 1999; Snyder and Borghard 2011). Nonetheless, analyzing the quality and strength of different proxies has proven difficult (Levy 2012; Slantchev 2012). Scholars have yet to directly test the hypothesis that democratic or authoritarian institutions covary with audience costs because we lack a sufficient model for their measurement independent of proxies relying heavily on regime-type information.

Second, the substantive effects of audience costs on conflict initiation are unclear due to two countervailing effects. On the one hand, a broad literature argues that increases in audience costs may discourage a given country from initiating disputes if the country expects its opponents to repeatedly stand firm in the future (Kurizaki and Whang 2015; Prins 2003; Weeks 2012). On the other hand, larger audience costs may also encourage their country to initiate a dispute, if they simultaneously coerce opponents to more quickly back down in the subsequent interaction (Schultz 1999). Thus, when countries internalize the long-term strategies of their opponents, their unobservable strategies create a time dependence between current escalation decisions and the expected path of future conflict, confounding the rela-

tionship between audience cost measures and dispute initiation. Such dynamic considerations can also effect the propensity for countries to back down. For example, a country may be willing to back down and incur a relatively large audience cost in a dispute today only if it expects tomorrow's resulting peace to be stable over the long run. In contrast, it would be less willing to incur those costs if the peace is merely transitory.

In this paper, we address these issues head-on by structurally estimating a dynamic game-theoretic model of crisis escalation. Using an explicit game form, we model audience costs as a parameter capturing the (dis)utility a country receives when it backs down from a dispute before its opponent (as in Fearon 1994*a*). The game is infinitely repeated, and countries fully anticipate the expected evolution of conflict and the possibility of incurring audience costs in equilibrium. We estimate audience costs using country-specific fixed effects, which do not depend on *a priori* determined variables, and we select the audience costs (along with other parameters and equilibria) that maximize the likelihood of the observed data. Specifically, we fit the model to Militarized Interstate Dispute Incident Profiles (MID-IP) data that record escalation decisions at the monthly level between 1993–2007 by using a new constrained maximum likelihood estimator developed by Su and Judd (2012) for dynamic models. Three major results emerge.

First, contrary to prevailing intuitions, we find that the second countervailing effect of audience costs dominates in the data. That is, increasing a given country's audience costs encourages it to initiate disputes along the path of play, an effect that emerges in 81% directed dyads. In the estimated equilibria, audience costs serve as a commitment device: they tie the hands of their respective countries, thereby encouraging them to stand-firm and coercing their rivals to back down in disputes. We find the hand-tying effect in 78% of directed dyads, where increasing a country's audience costs forces it to stand firm more frequently. Likewise, the coercion effect appears in 75% of directed dyads, where increasing a country's costs results in its opponent conceding more often. For highly related reasons, we find that increasing countries' audience costs discourage their opponents from initiating a dispute, an effect which emerges

in 84% of directed dyads. Although the hand-tying and coercion effects are well-studied in the literature, the finding that audience costs incentivize countries to initiate conflicts runs counter to previous finding in reduced form analyses (Clark and Nordstrom 2005; Prins 2003; Weeks 2008). In addition, it represents the first empirical support for the predictions in Schultz (1999). While audience costs and conflict initiation are negatively associated in some of our estimated equilibria, this trend appears in only a small minority of directed dyads (19%).

Second, we test the hypothesis that standard proxies for audience costs correlate with our estimates and find that the proxies are fair, but underwhelming, first approximations. Although democracy and authoritarian regime types are important predictors of our estimated audience costs, other systemic and domestic factors influence their magnitude. For example, the existence of an interstate rival attenuates the penalties leaders face for backing down from a dispute. Democracies with rivals have, on average, audience costs that are roughly similar to autocratic regimes with legal provisions for executive removal (e.g., China), suggesting that democratic voters may provide leaders some leeway when they escalate disputes with rivals. Additionally, a free press can strongly increase audience costs, confirming results derived from the formal model in Slantchev (2006).

Third, we use the fitted model to address a puzzle raised in the previous discussion: do higher audience costs lead to more or less conflict worldwide? On the one had, increasing a country's audience costs incentivizes that country to stand-firm and initiate disputes, leading to shorter peace and longer crises. On the other hand, audience costs also coerce rivals to back down and refrain from dispute initiation, which would increase the propensity for peace. We find that the latter effects dominate in the data. On average, audience costs are *peace enhancing*, that is, in two-thirds of directed dyads, larger costs reduce the long run probability that the dyad enters a dispute. Similarly, in 80% of undirected dyads, an increase in both countries' audience costs results in a higher propensity for peace in the long run.

Finally, while the analysis primarily contributes to debates concerning audience costs, our

structural framework is rich enough to produce several additional substantive implications for the wider international relations literature. Specifically, we find that trading partners and joint democracies are less inclined to enter wars, while no such aversion to crises exists. Thus, a Kantian or liberal peace may prevent more hostile conflicts, but not lower-level disputes. In addition, our results provide mixed support for the two systemic theories of conflict in Braumoeller (2008). In peace, the expectation of conflict deters escalation, but in crisis and war, it encourages or spirals further escalation. Such nuances offer one explanation as to why conflicts cluster temporally and peace is self-enforcing (Beck, Katz and Tucker 1998; Gleditsch and Ward 2000).

2 Modeling Audience Costs and Disputes

Before presenting formalities, we discuss the substantive motivation behind our approach to modeling audience costs and how it relates to the wider literature. Our first goal involves adopting a tractable definition of audience costs that permits identification from standard event data on interstate disputes. To do this, we follow the lead of Fearon (1994*a*) and define audience costs as the (dis)utility a country receives from backing down in an international dispute before its opponent.[1] In a similar vein, Slantchev (2012) defines the first core premise of audience cost theory as "backing down in a crisis makes [a country] suffer costs in addition to those arising from conceding the contests..."(p. 377).

Although this conceptualization holds a prominent place in the literature, it complicates and side steps two theoretical explorations that have developed since Fearon's seminal work. In one development, scholars often interpret audience costs as the punishment leaders receive from initiating a threat but failing to follow through on it.[2] In this case, the dispute's history

[1] "If [country] i quites the crisis before the other has quit or attacked, then... i suffers audience costs..." (Fearon 1994*a*, p. 582).

[2] In the experimental literature, audience costs refer to the disapproval leaders create when then say one thing but do another (Tomz 2007).

determines which countries receive audience costs, and the costs only affect the initiator's payoffs (Kurizaki and Whang 2015; Schultz 1998, 1999; Weeks 2008). This approach is not dominant as other theories require both initiators and targets to receive costs for backing down, however (Fearon 1994*b*; Kurizaki 2007; Schultz 2001). Such a setup is appropriate when "crisis diplomacy takes place before domestic audiences on both sides" (Kurizaki 2007, p. 545). In the second development, researchers often endogenize the magnitude of audience costs, allowing them to grow over the duration of the dispute or be products of strategic choices (Fearon 1994*a*; Leventoglu and Tarar 2005). Nonetheless, all previous empirical work, including both reduced form and structural approaches, treat audience costs as fixed and not being subject to strategic choices. In this paper, we adopt a more simplified conceptualization, where audience costs are country specific and do not depend on the historical evolution of disputes. This helps simplify already complicated estimation and counterfactual exercises, thereby serving as a useful starting point to a problem in which theory has traditionally outpaced empirics.

Our second goal captures the idea that countries are long-lived and decide whether to engage in conflict given their opponent's expected actions not only today but also tomorrow. This is important for several reasons. Long-standing theories, historical accounts, and common intuitions maintain that countries and national leaders are strategic and forward-looking. Thus, long-term expectations influence whether countries incur audience costs. In a crisis, a country may be more likely to back down and incur an audience cost today if it expects a stable peace to emerge subsequently. In this case, a country trades an immediate (audience) cost for a delayed benefit (peace). Conversely, it would be less likely to back down when the benefits of peace are more fleeting. If the analysis ignores these dynamic costs and benefits, then our audiences cost estimates will be misleading and difficult to interpret. Thus, we consider an interaction that is infinitely repeated, which precludes using off-the-shelf one-shot models as in Lewis and Schultz (2003). Although the proposed model closely resembles Fearon (1994*a*), we incorporate two notable differences. Strategic interaction does not end after backing down or

war, and countries endogenously begin disputes, essentially allowing for an infinite interaction.

Third, our framework remains agnostic about the particular mechanisms generating audience costs, but we investigate their determinants in a post-estimation exercise. This approach has several advantages. Most prominently, by using fixed effects, we avoid introducing an avenue of omitted variable bias into the analysis, which is likely to arise because audience costs originate from several highly correlated factors including democratic institutions, voter repressiveness, leader removal, national honor, among others (Chiozza and Goemans 2011; Dafoe and Caughey 2016; Levy 2012; Smith 1998; Tarar and Leventoğlu 2009). If we include some but not all of these factors, then our predicted audience costs and inferences concerning their origins would be biased.[3] In a similar vein, fixed-effects reduce the potential for issues, such as separation, that would arise from using several of these highly collinear variables.

Because we adopt a structural approach, our endeavor is most similar to Kurizaki and Whang (2015), and we build upon their work in four ways. First, they proxy audience costs using polity2 by assuming audience costs are a linear function of democracy; we impose no such assumption. Second, the two theoretical models differ considerably. Kurizaki and Whang (2015) use a version of the one-shot crisis-signaling model with sequential moves from Lewis and Schultz (2003), in which initiators only incur audience costs once. In this paper, we construct a dynamic model with simultaneous choices, where countries are infinitely lived, accommodating the long-term costs and benefits of backing-down and standing firm. Third, a draw back from using the crisis-signaling model to study interstate conflict is that the model almost certainly requires a very specific dataset from Schultz, Lewis and Zucco (2012), which covers the inter-war period (1919–1939).[4] In contrast, the dynamic model is more flexible with its informational requirements, and we use the standard MID-IP dataset, covering a

[3]An artifact of this approach is that we do not directly compare our audience cost estimates to their substantive effects on leader approval identified in the experimental literature (Levy et al. 2015; Tomz 2007). In a post-estimation exercise, we analyze the relationship between our estimates and other factors when relevant measures exist across countries. Nonetheless, even if we had measures of voter responsiveness to backing down across countries, public opinion data would present difficulties due to additional confounders.

[4]For a broader discussion of data issues in the crisis-signaling model, see the concluding paragraph from Lewis and Schultz (2003).

more temporary period (1993–2007). Finally, although the two models are quite different, both may admit multiple equilibria under certain payoffs. Standard estimation techniques, e.g., Signorino (1999), do not account for this multiplicity, leading to inconsistent estimates and incorrect counterfactuals (Jo 2011). Our estimation strategy and counterfactual exercises, however, avoid these issues.

3 Structural Model

This section presents the dynamic model of crisis escalation, which we subsequently fit to data. Because estimation of the game is our end goal, we include action-specific shocks that are private information and allow payoffs to depend on observed covariates. Previous work often presents additional versions that exclude these shocks and covariates (e.g., Carter 2010; Gent 2007; Kurizaki and Whang 2015). In contrast, we present a unified model upfront for two reasons. First, the unified model with shocks fits the data quite well, and a pure strategy equilibrium from a game without shocks would not have enough leverage to rationalize the patterns in the data. Second, with an infinite horizon, it is more involved to characterize equilibria without shocks, because there are multiple equilibria under some parameter settings and we cannot use backward induction as in earlier papers. Our presentation of the unified model spares readers from an additional section of rather involved formalities. Nonetheless, Appendices A and B contain additional information on our equilibrium characterization and a numerical example illustrating comparative statics, respectively.

Consider two countries. We use i to denote an arbitrary country and $j \neq i$ its opponent. Time is discrete and indexed by $t = 1, 2, \ldots$. In each period t, country i first observes a common state variable $s^t \in \{1, 2, 3\}$ and a private state variable ε_i^t, which represents private information that country i has about the costs/benefits of taking particular actions in period t. This information takes the form of action-specific shocks and is unknown to its opponent. Here s^t denotes the current level of hostility, where $s^t = 1$ denotes that the countries are in a state

of peace, $s^t = 2$ a state of crisis, and $s^t = 3$ a state of war.[5] Each country then simultaneously chooses a level of hostility against its competitor. Let $a_i^t \in \{1, 2, 3\}$ denote country i's action in period t, and a profile of actions is $a^t = (a_i^t, a_j^t)$. Here, a_i^t takes the values 1, 2, and 3 which indicate peaceful, crisis-level (threat/demand), and war-level (attack/invasion) actions, respectively.

The common state variable s^t evolves according to past actions, and we assume escalation is deterministic and unilateral, that is, $s^t = \max\{a_i^{t-1}, a_j^{t-1}\}$.[6] Thus, the model captures situations in which a country declares war ($a_i^t = 3$) on its opponent, and the next period begins with the two countries in a state of war ($s_i^{t+1} = 3$). We denote country i's private information about action a_i in period t as $\varepsilon_i^t(a_i^t)$. The private information, $\varepsilon_i^t(a_i^t)$, is independently and identically distributed type I extreme value across actions, players, and states, which are standard distribution and independence assumptions in these types of games.

Let θ denote a vector of relevant structural parameters to be estimated. Country i's per-period payoff against country j is given as $u_{ij}\left(a^t, s^t; \theta\right) + \varepsilon_i^t(a_i)$, where u_{ij} is i's deterministic utility and ε_i is i's (stochastic) private information. Given a sequence of action profiles, states, and action-specific shocks $\{(a^t, s^t, \varepsilon_i^t)\}_{t=1}^{\infty}$, country i's total payoff is the discounted sum of per-period utilities:

$$\sum_{t=1}^{\infty} \delta^{t-1} \left[u_{ij}\left(a^t, s^t; \theta\right) + \varepsilon_i^t(a_i^t)\right],$$

where $\delta \in [0, 1)$ denotes a common discount factor.[7]

[5]Hereafter, a *state* denotes the commonly observed level of hostility s^t, and we refer to the game's actors as countries. Additionally, we use the terms "dispute" and "conflict" interchangeably to refer to periods in which the path of play resides in states 2 and 3.

[6]Unilateral escalation is common in the crisis and conflict literature (Bueno de Mesquita, Morrow and Zorick 1997; Fearon 1994*a*; Kurizaki and Whang 2015; Schultz 2001).

[7]Notice we do not include δ in the parameters to be estimated. Without additional structure, it is difficult to identify δ in these types of models, and previous papers traditionally assume a fixed discount factor throughout (Arcidiacono et al. 2016; Magnac and Thesmar 2002; Pakes, Ostrovsky and Berry 2007). Here, we fix $\delta = 0.9$ as in Pesendorfer and Schmidt-Dengler (2003).

We endow u_{ij} with the following functional form:

$$u_{ij}(a, s; \theta) = \underbrace{x_{ij} \cdot \beta(s)}_{\substack{\textbf{state-specific} \\ \textbf{payoff}}} + \underbrace{z_i \cdot \kappa(a_i)}_{\substack{\textbf{action-specific} \\ \textbf{payoff}}} + \underbrace{\alpha_i \mathbb{I}\left[a_j \geq s > a_i\right]}_{\substack{\textbf{country-specific} \\ \textbf{audience cost}}} + \underbrace{\gamma(s) \mathbb{I}\left[a_i > 1\right] \mathbb{I}\left[a_j > 1\right]}_{\substack{\textbf{spiral/deterrence} \\ \textbf{effect}}}. \quad (1)$$

Country i's utility consists of four components. First, it receives a state-specific payoff, $x_{ij} \cdot \beta(s)$, from being in state s with country j, where x_{ij} is a vector of dyad-specific variables and $\beta(s)$ a vector of associated coefficients. Dyadic variables could be directed, e.g., military capability ratios, or undirected, e.g., minimum democracy.[8]

Second, regardless of the state, if country i chooses action a_i, i pays some costs $z_i \cdot \kappa(a_i)$, where z_i is a vector of country-specific variables and $\kappa(a_i)$ is a vector of associated coefficients. These costs of escalation capture important transaction costs from declaring war, formally threatening a opponent, or maneuvering military troops to a boarder area.[9] Notice that i's cost of action a_i does not depend on characteristics of j. This independence is an important identification assumption, but paired with the state-specific payoff, this leads to a natural interpretation: while the U.S. pays the same cost from declaring war on Afghanistan and on Russia, it can still possess a preference for being at war with Afghanistan over being at war with Russia.[10]

Although we operationalize the costs of the crisis- and war-level actions in a similar manner, the model incorporates the possibility that these actions entail more substantive differences. The two actions transition the game to strategically different states, which entail different payoffs. In addition, depending on the current state, the two actions generally produce diverging propensities for incurring audience costs. This setup, albeit relatively simple, is rich

[8]The payoff $x_{ij} \cdot \beta(s)$ may represent country i's expected utility from some lottery or game as long as the outcome does not affect payoffs or transitions. In the classic war lottery where p_{ij} represents a probability of victory, π_{ij} the benefit of winning, and c_{ij} the cost of war, the war-state payoff takes the form $x_{ij} \cdot \beta(3) = p_{ij}\pi_{ij} - c_{ij}$.

[9]This may appear to be a non-standard modeling choice because these costs of escalation are not considered in previous models, but this version subsumes the case in which the coefficients $\kappa(a_i)$ are zero.

[10]Without loss of generality, we adopt the normalization that $\beta(1) = \kappa(1) = 0$. That is, the payoffs $x_{ij} \cdot \beta(s)$ and $z_i \cdot \kappa(a_i)$ are relative to the baseline payoffs for the peaceful state and action, respectively.

enough to uncover empirical differences once the model is taken to data. We find that the propensities for choosing the crisis- and war-level actions vary across the three states and that the signs and significance of several coefficients vary across the crisis and war states/actions.

The last two components in a country's utility function depend on the actions of its opponent. The parameter α_i is a country-specific value and measures i's audience costs. That is, if i and j are engaged in a dispute ($s > 1$), j continues or escalates the current level of conflict ($a_j \geq s$) and i backs down ($a_i < s$), then i incurs cost α_i. Intuitively, this means that countries never acquire audience costs when in peace or if their opponents do not escalate/maintain the current conflict. Essentially, countries still receive audience costs when they deescalate the dispute before their rival, although size of the audience costs are fixed throughout the dispute.[11] The $\gamma(s)$ parameters are state-specific values measuring how i's cost of escalation varies with j's actions in state s. When $\gamma(s) > 0$, i's cost of escalation ($a_i > 1$) decreases when j escalates in state s. Similarly, when $\gamma(s) < 0$, i's cost of escalation increases when j escalates.[12] Thus, $\gamma(s)$ represents other strategic incentives as to why a country does not escalate a conflict independently of audiences costs including potential second-strike (dis)advantages. For example, if i receives a large benefit in state s when its opponents attacks first, such as support from an international community, then $\gamma(s)$ would be negative.

We characterize Bayesian-Nash equilibria in stationary Markovian strategies (equilibria, hereafter) as is standard in these games. Consider i's net-of-shock expected utility from choosing action a_i in state s, denoted $v_i(a_i, s)$, and let v_i denote the vector of expected utilities for every country i. Because ε_i is distributed type 1 extreme value, i chooses a_i in state s with

[11] We could alter the model so that only initiators incur audience costs by expanding the state space, where there are four dispute states including crisis initiated by i and war initiated by i, for all countries i. Such an extension is non-trivial as it involves estimating an additional $2,148$ parameters describing equilibrium play. Instead, we allow both countries to incur audience costs in disputes as in Fearon (2004), Kurizaki (2007), and Schultz (2001).

[12] It is possible these mutual payoffs $\gamma(s)$ are actually realized in the next period rather than the current period. If this is the case, we can always weight the estimated parameters by δ^{-1}.

probability

$$P(a_i, s; v_i) = \frac{\exp(v_i(a_i, s))}{\sum_{a_i'} \exp(v_i(a_i', s))}, \tag{2}$$

that is, equilibrium choice probabilities take the standard multinomial logit form. As in Signorino (1999), the expected utilities, v_i, are endogenous to equilibrium play. Unlike most previous work, a closed-form solution for these values does not exist due to game's infinite horizon and simultaneous moves.[13] In Appendix A, we demonstrate that profile $v = (v_i, v_j)$ is an equilibrium if and only if it satisfies a system of 18 smooth equations, $\Phi(v; \theta) = v$.

With this characterization, it is possible to analyze comparative statics even before fitting the model to data, as in Carter (2010) and Gent (2007). Given the setup of our model, however, such an exercise may not be particularly informative for several reasons. First, there may exist multiple equilibria, that is, multiple solutions to $\Phi(v; \theta) = v$ for a fixed θ, a complexity that does not arise in models with finite time horizons and sequential moves. With this multiplicity, different equilibria will exhibit diverging comparative statics.[14] Second, the size and direction of the comparative statics are highly dependent on parameter values, many of which do not have theoretically informed magnitudes or even directions. Even when theoretical expectations exist, i.e., democracy levels and audience costs, a goal of the paper is to examine the degree to which these relationships are present in the data.

Nonetheless, we characterize an equilibrium for specific parameter settings in Appendix B to better illustrate the main dynamics in our model. Of interest, the constructed equilibrium demonstrates a novel comparative static: increasing a country's audience cost parameter increases the probability that we observe the given country's probability initiate a dispute along the path of play. Although Schultz (1999) has a similar result in a signaling model of

[13] McKelvey and Palfrey (1995) and Jo (2011) discuss similar issues.

[14] Models with multiple equilibria are not necessarily defective as they accommodate the possibility that norms or focal points are equilibrium selection devices. They present added difficulties during the estimation and counterfactual process, however.

crisis escalation, the comparative static is still emerges here without signaling incentives and in a setting where both countries in a dispute may incur audience costs. One reason for this relationship is that the country with enhanced (more negative) audience costs is more likely to stand firm conditional on being in a crisis. Furthermore, the country is less likely to back down and receive audience costs along the path of play. In the subsequent sections, we the fit the model to data and examine whether this relationship holds given our estimated equilibria and structural parameters.

4 Empirical Strategy

4.1 Constrained Maximum Likelihood

We estimate the model using a full-information constrained maximum likelihood estimator (CMLE), as advocated by Su and Judd (2012). Given our application, this estimator has significant advantages over previous methods (Aguirregabiria and Mira 2007; Hotz and Miller 1993; Rust 1987). The procedure does not repeatedly compute equilibria, a process that is further complicated by the possibility of multiple equilibria. In addition, it does not require consistent first-stages estimates of choice probabilities, which is particularly important for the rare-event nature of interstate disputes. Finally, the CMLE avoids convergence issues that arise when iterating two-step approaches (Egesdal, Lai and Su 2013).[15]

We consider D dyads or games as described above. We index dyads by $k \in \{1, ..., D\}$ and include the superscript k hereafter. We use data that can be summarized as a list $\{X, Z, Y\}$. Here X and Z are matrices of ordered-dyad and country-specific variables, respectively, which enter the stage utilities through Equation 1. In addition, Y is a collection of matrices, detailing observed state and action profiles for each dyad, i.e., $Y^k = \left(s^{kt}, a_{i^k}^{kt}, a_{j^k}^{kt}\right)_{t=1}^{T}$ and T is the total

[15] One drawback of the CMLE is that the procedure requires solving a constrained optimization problem, which may be difficult to implement using standard statistical software. Nonetheless, Su and Judd (2012) describe how researchers can use open-sourced, industrial software instead. Appendix C describes our implementation.

number of observed time periods. Let $\bar{\theta}$ denote the true vector of parameters. For each dyad k, we assume the data Y^k were generated from a *single* equilibrium, $\bar{v}^k$, i.e., $\Phi^k(\bar{v}^k;\bar{\theta}) = \bar{v}^k$. While multiple equilibria potentially exist in the game between the countries i^k and j^k, the procedure requires that Y^k comes from only one of these. Let $\mathbf{v} = (v^1, ..., v^D)$ denote the vector of all profiles of expected utilities. The log-likelihood takes the following form:

$$\mathcal{L}(\mathbf{v} \mid Y) = \sum_{k=1}^{D} \sum_{t=1}^{T} \left[\log P\left(a_{i^k}^{kt}, s^{kt}; v_{i^k}^k\right) + \log P\left(a_{j^k}^{kt}, s^{kt}; v_{j^k}^k\right) \right], \tag{3}$$

which is a standard multinomial log-likelihood summed over dyads, time periods, and players. With a slight abuse of notation, the CMLE estimates, $(\hat{\mathbf{v}}; \hat{\theta})$, solve the following constrained optimization problem:

$$\begin{aligned} &\max_{(\mathbf{v};\theta)} \ \frac{1}{T} \mathcal{L}(\mathbf{v} \mid Y) \\ &\text{subject to } \Phi^k(v^k; \theta \mid X, Z) = v^k, \ k = 1, ..., D. \end{aligned} \tag{4}$$

Standard results on Lagrange multiplier tests, found in Silvey (1959), guarantee that the CMLE is consistent in T and characterize the estimator's asymptotic distribution. Consistency in the number of games or dyads is not guaranteed, as there is an obvious incidental parameters problem. Nonetheless, we gain leverage by pooling information across dyads when T is sufficiently large, and this pooling is necessary for identification with a larger number of variables. We relegate further estimation details to Appendix C, and Appendix D contains a Monte Carlo illustrating the properties of the CMLE on datasets of similar size to the one we construct in the subsequent section.

4.2 Data

We use the MIDs 4.0 data (Ghosn, Palmer and Bremer 2004) to define each dyad's observed path of play, Y^k. In particular, we make use of the incident level data known as MID-IP

4.01 (Kenwick et al. 2013). The data record actions taken by the individual countries within interstate disputes between 1993 and 2010. These actions are then used to create the state transitions. Dispute numbers determine what country or countries the actions were taken against. In our framework, a time period is a calender month, because approximately 50% of incident reports include no more precise timing information. The actions recorded by the MID-IP are on the standard 22-point MID scale, ranging from no action to joining an interstate war. We use this scale to form the three levels of hostility countries can take against each other: war, crisis, and peace. We code a 'war-level' action if country i attacks or takes a more hostile action against country j in period t (MID-IP actions 16-21). A 'crisis-level' is recorded if the country commits an action that is between a threat and an attack (MID-IP actions 1-15).

We follow Whang, McLean and Kuberski (2013) and construct the dataset in two steps. First, to avoid selection bias, we fill in peaceful actions for all country-dyad-months in which the MID-IP database does not include a military incident (Huth and Allee 2002). Second, we define a set of "politically relevant dyads" that restricts the sample to every dyad that has entered the MID-IP data. Ultimately, the data contain 179 dyads with 180 time periods each; approximately 95.4% of observed states are peace while 97% of actions are peace-level. Table 1 records the nine different types of possible transitions and provides preliminary evidence that countries condition their behavior on the state variable of interest. That is, the conditional distribution of transitions changes substantially across the observed states.

Using this data, the model gains leverage on estimating audience costs through two observable moments: the probability with which a country (a) initiates disputes in peace and (b) backs down within a dispute. To see this, hypothetically fix the strategies of the two actors in all periods such that they place positive probability on every action in every state. Then increasing i's audience costs (i.e., α_i moves to $-\infty$) has two important effects. First, i's expected utility of initiating conflict ($a_i > 1$) once in peace ($s = 1$) decreases, because initi-

Table 1: Distribution of transitions in the data.

Transition	Percent of data	Percent within each state
Peace → Peace	92.5%	97.0%
Peace → Crisis	1.82%	1.91%
Peace → War	1.09%	1.14%
Crisis → Peace	1.81%	71.8%
Crisis → Crisis	0.49%	19.6%
Crisis → War	0.22%	8.62%
War → Peace	1.09%	53.6%
War → Crisis	0.21%	10.2%
War → War	0.74%	36.2%

Caption: The middle column displays the probability distribution over the possible transitions, and the far-right column presents the conditional distribution in each state.

ation certainly transitions the game into a state in which i receives an audience cost with a fixed probability. Second, i's expected utility of playing the peace action ($a_i = 1$) in a dispute ($s > 1$) will change as well, but the direction of this change will depend on j's strategy. If j stands firm with a sufficiently large probability, then i's utility of playing peace will decrease as it expects to subsequently incur audience costs. In contrast, if j is likely to play the peaceful action, then i's utility of playing peace may increase as i would prefer to deescalate the dispute, transitioning the to a peaceful state without audience costs.

These dynamics have two important implications. First, we cannot identify a country's audience cost parameter if it has never been in a dispute with another country in the data, although we do not require that it initiates or backs down in a dispute. Including such a country, e.g., Costa Rica, would lead to separation because its contribution to the likelihood function will be strictly increasing as its audience cost parameters become more negative.[16] Second, the above discussion highlights the importance of equilibrium analysis. Essentially, the discussion fixes the strategies of both players, thereby ignoring the indirect effects of audience costs that will change j's equilibrium strategy and therefore i's expected utility

[16]In a similar vein, countries that only enter one dyad and one crisis within that dyad, e.g., Ghana in the South Africa–Ghana dyad, tend to have larger standard errors associated with their audience parameter.

calculations as well. As we demonstrate below, the estimated equilibria can and often times produce comparative statics that diverge from the naive predictions in the previous paragraph.

To isolate the effects of audience costs, we also control for other reasons why countries initiate a dispute or back down. First, we control for the possibility of second-strike (dis)advantages or a general preference for peaceful actions and states. The former are controlled for by the parameters $\gamma(s)$, while the latter are captured by including constants in x_{ij} and z_i. In addition, we include other control variables common to the interstate conflict literature. At the dyadic-level, we use the minimum democracy level in the dyad, the logged capability ratio, and the square-root of the trade interdependence. We measure the minimum level of democracy using the standard Polity IV database. Capability ratios are computed as the ratio of CINC scores from the Correlates of War (COW) National Material Capabilities 3.0 (NMC) dataset (Singer, Bremer and Stuckey 1972). Trade interdependence is measured in the usual fashion (Gartzke 1998; Oneal and Russett 1997), where country i's interdependence on country j is the sum of exports and imports between i and j divided by i's GDP. Trade data comes from the COW dyadic trade data (Barbieri, Keshk and Pollins 2009), supplemented by data from Gleditsch (2002). GDP data is from the Penn World Table (PWT) 8.0 (Feenstra, Inklaar and Timmer N.d.) and supplemented with data from the World Bank. For the variables associated with country specific costs, we include logged GDP per capita (from PWT) and logged military personnel per capita (from NMC).

We take the mean values over the course of the time period in the study (1993-2007) to produce values in x_{ij} and z_i. While there is a legitimate concern that some these measures are endogenous to the conflict process itself, the variables in the analysis show little change over the time frame considered here. Even when these variables do change, there is no correlation between these changes and the observed states and actions. See Appendix E for more details.

5 Audience Costs

Figure 1 presents the point estimates for the 125 country-specific audience costs and their 95% confidence intervals, sorted by magnitude.[17] Even though the parameters α_i can take on any value in estimation, all the point estimates are negative, suggesting that leaders are punished for backing down. Interestingly, the countries with the ten largest (most negative) audience costs are mostly democracies, but notable exceptions exist to the idea that democracy and audience costs are synonymous. Many autocracies and anocracies exhibit substantial audience costs, e.g., Turkmenistan and Belarus. This offers some *prima facie* evidence in favor of arguments suggesting that autocrats in weak states with real removal threats face large audience costs (Chiozza and Goemans 2011; Weeks 2008). Despite this, preliminary difference-in-means tests indicate that democracies have larger audience costs than both autocracies and anocracies ($p < .05$), but a test finds no difference between the latter two groups.

Although democracies have larger audience costs than autocracies on average, several exceptions to this trend are involved in historically salient and persistent conflicts. For example, North and South Korea have similar audience costs as South Korea's audience cost is very moderate (less negative) compared to other advanced democracies. One reason for this is that South Korea exists in a state of perpetual siege, suggesting that voters are more willing to give their leader free range to do whatever she thinks is best to avoid a costly war.[18] A similar story explains why Israel and India both have very low—for democracies—audience costs.

Finally, comparing across autocratic institutions, the audience cost parameters vary in expected ways. To see this, consider Weeks (2008, 2012). We analyze the degree to which our estimated audience costs match her theoretical predictions regarding autocratic regime

[17] Point estimates in table format can be found in Appendix F.

[18] Other explanations for South Korea's moderate audience costs include its dependence on U.S. military support and lack of Wartime Operational Control. The trend of smaller audience costs among democracies with rivals holds more generally, however. Audience costs and the number of rivals, as defined in Thompson and Dreyer (2012), have a correlation coefficient of 0.34 among democracies, which is significant at the $p < 0.01$ level.

Figure 1: Country-specific audience costs, labeled with three-letter COW codes.

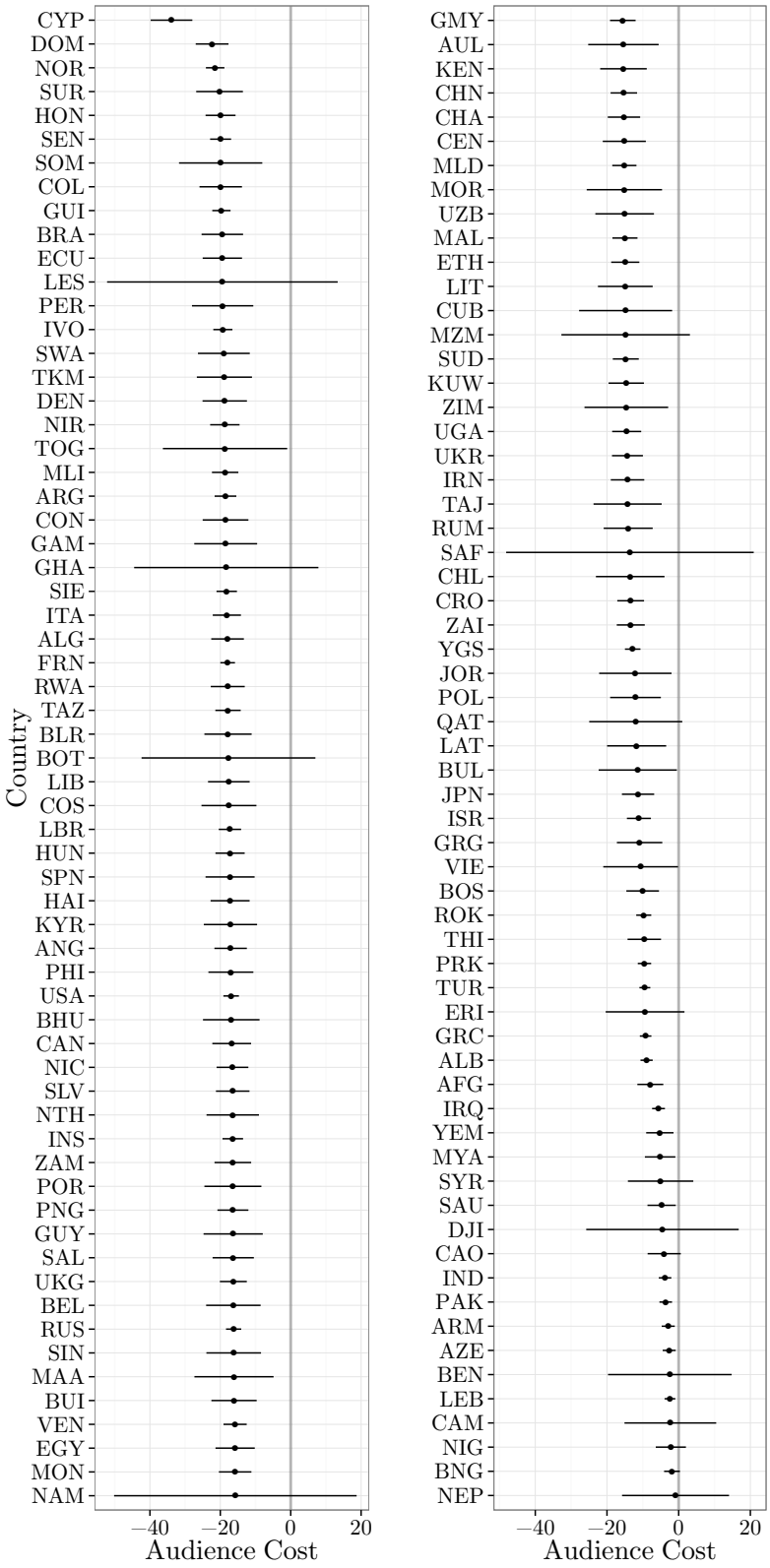


Table 2: Regime classification and audience costs

	WLS: Audience Costs	
	Weeks (2008)	Weeks (2012)
Personalist	4.89**	
	(1.67)	
Single-party	−1.11	
	(2.28)	
Military	2.39	
	(3.97)	
Machine		−1.43
		(1.63)
Junta		1.93
		(2.11)
Boss		3.94*
		(1.58)
Strongman		1.15
		(2.00)
Other non-democracy	1.69	2.13*
	(1.10)	(1.01)
Population	0.34	0.36
	(0.36)	(0.28)
Constant	−18.35**	−18.32**
	(3.80)	(2.85)
R^2	0.10	0.09
N	89	125

Notes: $^{**}p < 0.01$; $^{*}p < 0.05$

Standard errors in parenthesis. Observations are weighted by the number of dyads in which each country appears.

types.[19] In the case of Weeks (2008), we consider personalist, single-party, and military autocracies, while from Weeks (2012) we include machine, junta, boss, and strongman. In both cases, democracies are the excluded category. We uncover the trends she hypothesizes in Table 2, where personalist and boss leaders have smaller (closer to $+\infty$) audience costs than democracies. Furthermore, machines tend to have more intense audience costs than democracies, but the difference is not significant at conventional levels.

5.1 The Effects of Audience Costs

We consider how audience costs affect countries' propensity to (a) initiate disputes and (b) back down and actually receive audience costs along the path of play. Such an analysis allows us to describe the substantive effects of audience costs on observable behavior in interstate conflict. For each dyad, we compute the marginal effect of making audience costs more negative on each actor's equilibrium probability of backing down and initiating a conflict. We aggregate trends across individual dyads rather than constructing an "average" dyad, because there is information concerning what equilibrium such a dyad would play. Such an exercise is theoretical. It describes how the estimated equilibria change as functions of audience costs rather than correlating our audience cost estimates with the observed equilibrium choice probabilities in a reduced-form analysis. As discussed above, the latter approach produces misleading substantive effects as it ignores the indirect effects of audience costs through countries' equilibrium strategies.

Table 3 reports the percentage of directed dyads where the marginal effect of more intense audience costs on the indicated probability is positive. First, consider their effects on a country's likelihood of backing down and receiving an audience cost. In 78% of directed dyads, larger audience costs for country i decrease (increase) the frequency i backs down (stands firm) along the equilibrium path of play. This illustrate the direct, hand-tying effect

[19] Although only the former deals explicitly with audience costs, both generate predictions about the political costs leaders face from domestic audiences.

Table 3: Marginal effects of audience costs across dyads.

	Marginal Effect of $\alpha_i \to -\infty$
$\uparrow$ Pr(i Backs Down)	22%
$\uparrow$ Pr(j Backs Down)	75%
$\uparrow$ Pr(i Initiates)	81%
$\uparrow$ Pr(j Initiates)	16%
$\uparrow$ Pr(Peace)	65%

Caption: Percentages denote the proportion of directed dyads where increases in audience costs increase the probability of backing down, initiating conflict, and the long-term probability of peace. The probabilities, and their associated derivatives, are formally defined in Appendix B.

of audience cost, where a country is less likely to concede a dispute as its audience costs increase. Likewise, larger audience costs for country i increase (decrease) the probability its opponent j backs down (stands firm) along the path of play in 75% of directed dyads, This is an indirect effect of audience costs through j's strategy, matching the coercive effects discussed in previous theoretical and empirical work (Downs and Rocke 1995; Fearon 1994*a*; Kurizaki and Whang 2015; Partell and Palmer 1999; Uzonyi, Souva and Golder 2012).

Next, we examine the effects of audience costs on conflict initiation. Counter to intuition, the results in Table 3 show that, on average, audience costs embolden their leaders into initiating disputes with greater frequency. Specifically, in 81% of directed dyads, as i's audience costs increase, i is more likely to initiate a dispute along the path of play, *despite* the possibility that it might pay these larger costs later. This finding runs counter to arguments and empirics in other work, where authors find that larger audience costs temper the propensity for countries to risk conflict (Clark and Nordstrom 2005; Kurizaki and Whang 2015; Prins 2003; Schultz 1998; Weeks 2012).[20] Nonetheless, it provides some of the first empirical support for Schultz's (1999) prediction that increased audience costs result in more crisis onsets. Even though we find that audience costs can lead their respective countries to initiate less, this effect is not prominent in the data, arising in only 19% of directed dyads.

[20]Previous work covering the 1993–2007 time frame uses standard, reduced form regressions. Although Kurizaki and Whang (2015) adopt a structural approach, they use data from 1919–1939.

The emboldening effect arises from credible commitments. Essentially, when country i has larger audience costs, it becomes more likely to commit to standing firm in the subsequent dispute, as in the hand-tying effect described above. Furthermore, these larger audience costs coerce i's opponents to back down, as in the coercive effect described above. Countries internalize these advantages when deciding whether to initiate disputes. Because higher audience costs commit their respective countries to stand firm and coerce their rivals to back down, they also encourage countries to initiate disputes, as leaders attempt to exploit their enhanced credibility by becoming more aggressive in peaceful states.

Conversely, in 84% of dyads, j becomes less likely to initiate disputes as i's audience costs increase. This effect arises for the same credibility concerns described above. When i has larger audience costs, country j knows that its opponent will more easily stand firm during disputes, and, as a result, j is relatively disadvantaged and will more likely concede. Thus, j avoids this relative weakness within disputes by initiating less, that is, increasing i's audience costs deters opponent j beginning new disputes.

These results motivate the fundamental question: are larger audience costs, on average, beneficial for world peace? On the one hand, raising a country's audience costs encourages it to initiate conflicts and stand-firm in disputes. On the other hand, larger costs also encourage opponents to back down and not initiate new disputes. Which effect dominates in the data? To answer this, we calculate the marginal effect of making audience costs more negative on each dyad's probability of peace in the long run.[21] This effect is the rate of change in the incidence of peace as audience costs tend toward negative infinity. Simple tallying demonstrates, that in 65% of directed dyads, increasing an actor's audience cost results in more peace. Specifically, in 38% of dyads, increasing either country's audience costs makes both sides more peaceful, while both sides becomes more conflictual in only 8%. In the remaining 54%, the effects are mixed, where larger audience costs for i result in more peace but larger costs for j result in

[21]The long run probability of peace comes from the dyad's invariant (stationary) distribution. We describe this in Appendix B. In Appendix J, we provide evidence that equilibrium play has converged to its invariant distributions in most dyads.

less peace. Furthermore, when we sum the marginal effects of increasing both actors' audience costs, we find that in 80% of undirected dyads the total effect is an increase in peace. These substantive effects demonstrate that the dominant effects of audience costs is their deterrence value: countries tend not to initiate disputes against countries with higher audience costs, leading to more peace in the long run.

5.2 The Correlates of Audience Costs

We now explore the best proxies for audience costs by first considering how well standard proxies associate with the estimates in Figure 1. Table 4 reports the relevant correlation coefficients. In particular, we consider a country's polity2 score, Bueno de Mesquita et al.'s (2005) W,[22] a dummy for free press from Li (2005) and supplemented by Freedom House (Karlekar and Dunham 2012), whether the executive is directly elected (Regan, Frank and Clark 2009), whether there are constitutional provisions exist for executive removal (Regan, Frank and Clark 2009), a measure of executive constraints from the polity2 data, the number of interstate rivals the country has (Thompson and Dreyer 2012),[23] and a past proxy for audience costs from Uzonyi, Souva and Golder (2012), which they call "audience cost capacity" (ACC).

The simple bivariate relationships tell us that many of the standard proxies do a fair, but not impressive, job of capturing audience costs. The democracy based proxies are all significant and correlate in the expected direction. Because Kurizaki and Whang's (2015) audience cost measures are a linear function of polity2 scores, imputing audience costs using their coefficient estimates with polity2 scores from our data, i.e., mean polity2 between 1993–2007, will produce an correlation coefficient (0.18) identical in magnitude to that of polity2 in Table 4. Furthermore, the positive relationship between a free press and the strength of

[22]While not traditionally used as an audience cost proxy, winning coalition size (W) is frequently characterized as a stand-in for regime characteristics that are themselves associated with audience costs (Lektzian and Souva 2003; Weeks 2008).

[23]Thompson and Dreyer (2012) identify interstate rivals by "avoid[ing] the conflict record and focus[ing] on who state decision makers (or their historians) say are or have been their competitive and threating enemies" (p. 11).

Table 4: Correlates of audience costs.

	Spearman's ρ	t	p-value (one-sided)
Polity2	-0.18	-2.04	0.02
W	-0.13	-1.49	0.07
Free Press	-0.13	-1.43	0.08
Elected Executive	-0.20	-2.29	0.01
Executive Removal	-0.14	-1.51	0.07
Executive Constraints	-0.09	-1.04	0.15
Rivalry	0.32	3.78	0.00
USG's ACC	-0.25	-2.34	0.01

audience costs matches the theoretical predications in Slantchev (2006). These relationships, however, raise important questions about the appropriateness of using any one variable to stand-in for audience costs in empirical work.

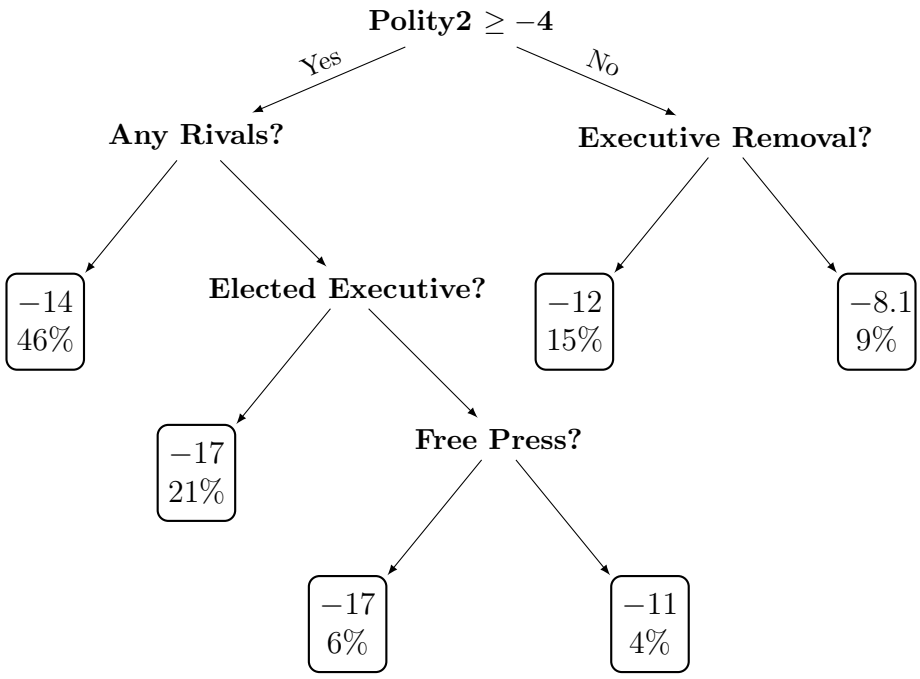
To analyze potentially more intricate relationships between these proxies and audience costs, we consider a regression tree. We include all the variables from Table 4, except for executive constraints and ACC, which are removed because they are composed of polity2 components. Additionally we add dummies for various types of democratic electoral systems from Regan, Frank and Clark (2009). The output of the regression tree is shown in Figure 2; values at each terminal node refer to the average audience cost among classified countries and the percentage of observations at the node.

According to this method, the best predictor of large audience costs is if a country has a polity2 score greater than or equal to –4.[24] Among autocracies, those that do not have any institutions for executive removal are the countries with the lowest audience costs. That is, the truest of autocrats have an average audience cost of –8 and include Saudi Arabia and Swaziland. In contrast, those with legal provisions for executive removal (e.g., China and Vietnam) have larger costs, with an average of –12.

Looking at democracies and anocracies, i.e., those with a polity score above −4, the

[24]This is similar to Kurizaki and Whang (2015), who find that a –5 polity2 score is an important cutoff. In their case, it is the cutpoint for whether audience costs exist. Here, it is the cutpoint that best predicts high versus low audience costs.

Figure 2: Regression-tree predictors of audience costs.



next most important predictor—before any other domestic institution—is the existence of an interstate rival. Countries in this group include many powerful states, such as the U.S. and Russia, as well as minor powers, such as Iran and India. Thus, while democratic institutions are generally associated with larger audience costs, having a rival attenuates their magnitude, an important caveat when considering the relationship between democracy and audience costs.

Countries with polity2 scores larger than –4 but without an interstate rival are split based on whether the chief executive is directly elected. In countries where there are direct elections, leaders face large costs with an average of –17, which is substantial compared to the other groups in the regression tree. Countries in this category include a grab-bag of strong and weak democracies, such as France, and Brazil. Notice that, in countries without direct elections of the executive, the existence of a free press has the same effect as direct election, i.e., a free press can generate relatively large audience costs, again confirming a result from the formal model in Slantchev (2006). Countries at this node are mostly parliamentary democracies.

The final node on the tree considers non-autocratic countries without rivals, a directly

elected executive, or a free press. These countries tend to be weak democracies and anocracies, such as Indonesia and Cambodia. Audience costs in these regimes are the lowest on this half of the regression tree, which indicate the necessity of certain domestic institutions, besides democracy levels, to constrain leaders. When non-autocratic regimes neither directly elect their chief executive nor have a free press, then they have audience costs with those that are comparable to autocrats. For example, Bangladesh with a polity score of six (until 2007) has an audience cost of -1.9, rivaling the least constrained of autocrats.

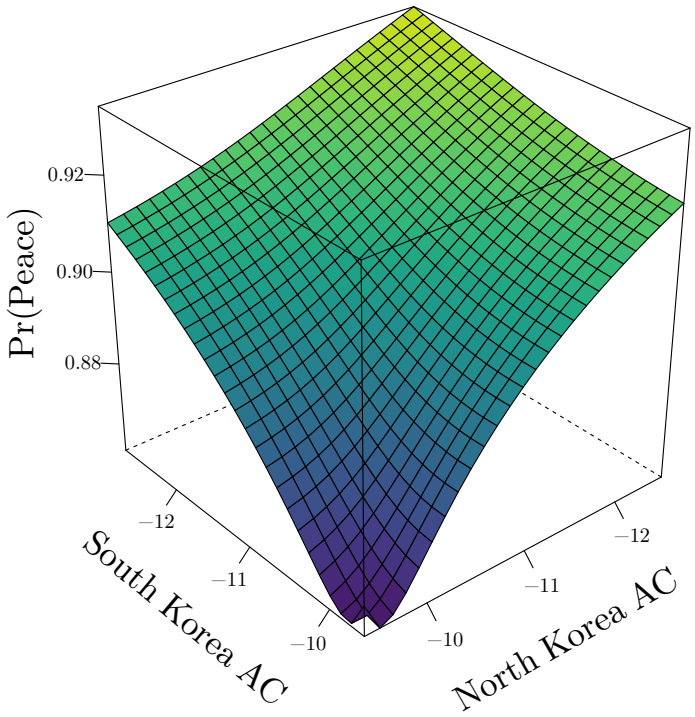
In sum, we find that democracy scores alone do not fully explain audience costs. Democrats have higher audience costs than autocrats, but conditioning on regime type, both domestic institutions within countries and rivalries among countries also affect leaders' costs of backing down. Strong democracies with rivals (e.g., South Korea and Israel) have audience costs that are effectively indistinguishable autocrats with legal executive removal (e.g., China). Thus, international factors are also important determinants of audience costs.

5.3 Illustrative Examples

To illustrate the findings in the previous sections, we examine two salient conflicts in greater detail. First, we consider substantive effects from the North and South Korea dyad. Figure 3 graphs the predicted probability of peace along the equilibrium path of play as a function of each player's audience cost.[25] The graph illustrates the pacifying effects of audience costs. As both costs become more severe (move towards $-\infty$), the likelihood of peace in the long run increases. A roughly two unit change in both cost parameters produces an eight percentage point increase in the probability that these countries are at peace.

[25]Multiple equilibria exist in each dyad, which complicates counterfactual experiments. Thus, we cannot just vary the parameters, compute a new equilibrium, and compare choice probabilities under the old and new parameters values. Doing so would not guarantee that the new equilibrium bears any resemblance to the estimated one. To address this problem, we use a method from Aguirregabiria (2012) that maps equilibria as locally continuous functions of data or parameters. Appendix G contains more details.

Figure 3: Audience costs and long-run peace: North-South Korea

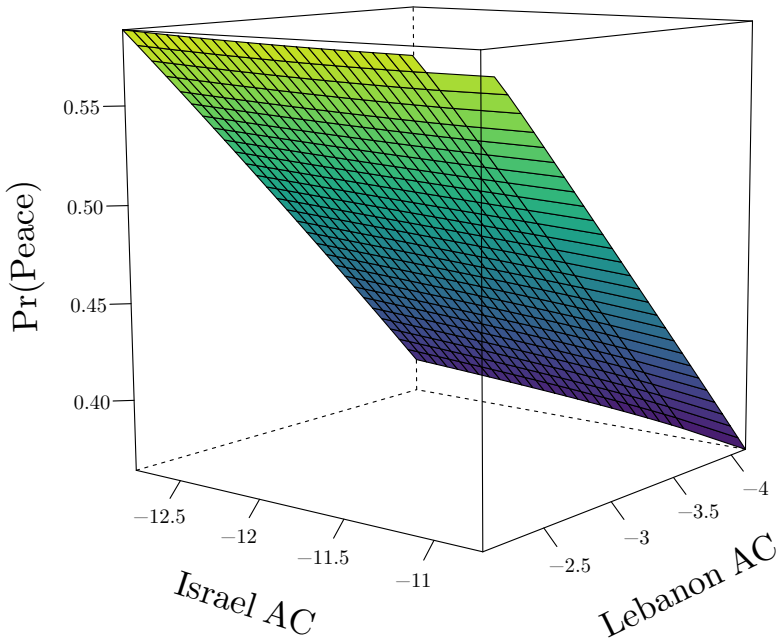


Some of these trends appear when considering how the relationship between the two countries evolved during 1990s and 2000s. After Kim Dae-jung assumes the presidency—representing the first peaceful transfer of power to an elected opposition party—South Korea finishes its democratic consolidation, and there is an unprecedented push towards peace, referred to as the Sunshine Policy (CNN 1998; Lee 2002). In this case, democratic consolidation in South Korea may have increased President Kim's audience costs.[26] Our analysis suggests that such a shift should make North Korea less likely to initiate a dispute, i.e., the coercive effects of audience costs. When this interstate rivalry re-solidifies in the early 2000s, we expect South Korea's audience cost to become less intense.[27] Here, our analysis suggests that North Korea becomes more likely to initiate a dispute as South Korea's audience costs move from –13 to –10, i.e., the range in Figure 3. In contrast, the probability of that South Korea initiates

[26]The substantive effects in Figure 1 could be interpreted as the effects of an unexpected and exogenous change in audience cost parameters on the equilibrium probability of peace. We maintain this interpretation throughout this section.

[27]Notably, President Bush's "axis of evil" speech is frequently credited with helping return this rivalry to saliency (Lee 2002; Paik 2002).

Figure 4: Audience costs and long-run peace: Israel–Lebanon



remains effectively unchanged. Matching this assessment, from the mid-2000s onward there have been a rash of North Korea initiations (Lavelle 2015).

In another case, consider the effects of audience costs on the conflict between Lebanon and Israel, shown in Figure 4. Here, increasing Israel's audience costs results in more peace but increasing Lebanon's produces more conflict. To see this, notice that if we were to fix Lebanon's audience cost and only increase Israel's by one unit, the dyad becomes more peaceful, where the probability of peace increases by approximately 2 percentage points. A point of interest is that Israel's 2006 invasion of Southern Lebanon occurred during a period where Wolf (2016) argues that the Israeli parliament was particularly fractured and unable to inflict audience costs on the prime minister (p. 431). In this case, the Israeli leadership faced lower audience costs due to poor coordination among opposition parties. Our analysis predicts that moving Israel's audience costs toward zero should decrease the likelihood of peace, matching the onset of the 2006 conflict.

In contrast, small increases in Lebanon's audience costs generate more conflict, where a unit increase (toward $-\infty$) results in a 10 percentage point decrease in the predicted probability of

peace. This comparative static is an example where audience costs are positively correlated with conflict in certain directed dyads. Although this effect does not dominate in the data—see Table 3—it appears in 35% of directed dyads. Upon further inspection, we find that larger audience costs for Lebanon discourage both sides from initiating disputes, *but* the disputes that do erupt, endure for longer periods. Specifically, adjusting Lebanon's audience cost over the range listed in Figure 4 increases conflict duration by three to four months on average.

6 Additional Results

In this section, we report the coefficient estimates associated with the independent variables that determine a country's state- and action-specific payoffs. Table 5 reports the results. The first two columns describes the estimates concerning state-specific payoffs that country i gets from its current state of affairs with country j. As with the standard multinomial logit, these estimates are interpreted as the relative increase or decrease in utility compared to being in the peaceful state. Columns three and four show our estimates of the country-specific costs to country i from taking action a_i. The last column contains the estimates of the structural parameters $\gamma(s)$ which describe whether conflict deters or spirals in state s.

There are three main things to note in this table. First we find a liberal peace in war but not crisis.[28] Specifically, the estimates suggest that as joint democracy levels and trade dependence increase, the relative benefit from being in the war state, compared to peace, decreases. This is in line with classic works like Oneal and Russett (1997) and Gartzke (1998). However, we find little support that democracies and trading partners avoid entering crises with each other.

Second, we see that action-specific cost estimates, $\kappa(a_i)$ have large negative constants. This is reassuring and provides some face validity to our results, because negative constants

[28] Appendix I contains additional substantive effects including the effect on joint democracy levels on the probability of peace.

Table 5: Structural estimates, omitting α_i.

	β(Crisis)	β(War)	κ(Crisis)	κ(War)	γ
Joint Democracy	0.00	-0.03^*			
	(0.01)	(0.01)			
Cap. Ratio	0.00	-0.05^*			
	(0.02)	(0.02)			
Trade Depend.	0.19	-3.06^*			
	(0.49)	(0.79)			
GDP pc			0.13^*	-0.15^*	
			(0.03)	(0.03)	
Mil. Per. pc			7.54	-8.15	
			(4.39)	(5.15)	
γ(Peace)					-48.24^*
					(2.65)
γ(Crisis)					9.32^*
					(0.60)
γ(War)					13.84^*
					(0.59)
Constant	19.23^*	19.08^*	-21.18^*	-20.96^*	
	(1.10)	(1.40)	(1.02)	(1.27)	
Log L			-45.50		
Dyads			179		

Notes:$^*p < 0.05$; Standard errors in parenthesis

demonstrate that the baseline cost of starting a crisis or conflict is indeed utility reducing. Increasing either GDP per capita decreases the cost of playing the crisis action. This demonstrates that wealthy and militarily powerful countries better afford crisis, and they pay less for these actions, an intuitively satisfying result.

Third, we consider the spiraling/deterrence effects of conflict (the $\gamma(s)$ coefficients) from the systemic theories of conflict in Braumoeller (2008). Note that $\gamma(\textsc{Peace}) < 0$, which suggests that, in peace, countries are playing something like chicken, and the expectation of conflict deters reciprocation because i receives a strong additional dis-utility from escalation when it expects j to escalate as well. In contrast, the positive coefficients on $\gamma(\textsc{Crisis})$ and $\gamma(\textsc{War})$ indicate that once countries are in a dispute, a country's cost of continuing the conflict decreases when its opponent continues as well. In these states, conflict spirals. This indicates that concerns about the security dilemma are particularly relevant when there is an active dispute between two countries.

Overall, these results provide the model with face validity, add an important qualification to liberal peace theories, and have several implications for systemic theories of conflict. Furthermore, we demonstrate that they are robust to several model specifications contained in Appendix H. For example, we control for other variables such as alliances, distance between two countries, and population. We estimate the model with one audience cost parameter, i.e., $\alpha_i = \alpha$ for all countries, and the substantive conclusions remain in tact. Finally, Appendix J contains model-fit exercises. We analyze the degree to which the model reproduces the predicted number of states and transitions both in aggregate and within each dyad. Several goodness-of-fit measures indicate good fit, and the structural estimator fits the data better than standard multinomial logits.

7 Conclusion

Audience costs are ubiquitous in theories of international relations. We adopt a structural approach to more flexibly estimate audience costs, examine their substantive effects, and test the strength of potential proxies. Our two major results concern the effects of audience costs on the evolution of interstate disputes. First, audience costs have an emboldening, not restraining, effect on conflict initiation, i.e., raising a country's audience costs makes it more likely to initiate disputes. This effect emerges because audience costs serve as a commitment device: increasing a given country's audience costs encourage it stand-firm during disputes, while coercing its opponents to more readily back down. When deciding whether or not to initiate, countries anticipate these effects that audience costs generate, so countries with larger audience costs may more freely initiate disputes in order to capitalize on their additional credibility. Second, despite this emboldening effect on conflict initiation, audience costs are, on the whole, peace enhancing. This result emerges from the deterrence effect of audience costs where raising one country's audience costs makes its opponents less likely to initiate disputes. The deterrent effect emerges appears because both sides anticipate the coercive and hand-tying effects: countries attempt to avoid initiating disputes with opponents who have larger audience costs and the additional leverage they entail.

In addition, we find that democracy explains only a portion of the variation in audience costs, and other domestic factors such as executive removal, directly elected executive, and a free press also contribute to audience costs independently. Our results also highlight an important international determinant: democracies with rivals tend to have less severe audience costs than democracies without rivals.

Our theoretically and empirically unified approach to analyzing conflict dynamics offers important avenues for future research. First, the model and estimator are flexible enough to incorporate other actors or a more nuanced state space. For the former, the model could be expanded to include international organizations such as NATO, and for the latter, we

could model the presence of nuclear weapons or diplomatic connections with additional state variables. Second, we provide flexible estimates of audience costs that can be used in future empirical studies as key variables of interest. Third, future work should consider structural models that allow audience costs to be strategically chosen or accumulated over the course of a dispute. Although such an endeavorer entails additional complications, it would better connect more modern theories of audience costs to the structural and empirical enterprise.

References

Aguirregabiria, Victor. 2012. "A Method for Implementing Counterfactual Experiments in Models with Multiple Equilibria." *Economics Letters* 114(2):190–194.

Aguirregabiria, Victor and Pedro Mira. 2007. "Sequential Estimation of Dynamic Discrete Games." *Econometrica* 75(1):1–53.

Arcidiacono, Peter, Patrick Bayer, Jason R. Blevins and Paul B. Ellickson. 2016. "Estimation of Dynamic Discrete Choice Models in Continuous Time with an Application to Retail Competition." *Review of Economic Studies* 83(3):889–931.

Barbieri, Katherine, Omar M.G. Keshk and Brian Pollins. 2009. "TRADING DATA: Evaluating our Assumptions and Coding Rules." *Conflict Management and Peace Science* 26(5):471–491.

Beck, Nathaniel, Jonathan Katz and Richard Tucker. 1998. "Taking Time Seriously: Time-Series-Cross-Sectional Analysis with a Binary Dependent Variable." *American Journal of Political Science* 42(4):1260–1288.

Braumoeller, Bear F. 2008. "Systemic Politics and the Origins of Great Power Conflict." *American Political Science Review* 102(1):77–93.

Bueno de Mesquita, Bruce, Alastair Smith, Randolph M. Siverson and James D. Morrow. 2005. *The Logic of Political Survival*. Cambridge, MA: MIT Press.

Bueno de Mesquita, Bruce, James D. Morrow and Ethan R. Zorick. 1997. "Capabilities, Perception, and Escalation." *American Political Science Review* 91(1):15–27.

Carter, David B. 2010. "The Strategy of Territorial Conflict." *American Journal of Political Science* 54(4):969–987.

Chiozza, Giacomo and H.E. Goemans. 2011. *Leaders and International Conflict*. New York: Cambridge University Press.

Clark, David H. and Timothy Nordstrom. 2005. "Democratic Variants and Democratic Variance: How Domestic Constraints Shape Interstate Conflict." *Journal of Politics* 67(1):250–270.

CNN. 1998. "Opposition boycott shadows South Korea's new president." *CNN World News*. Retrived from: `http://www.cnn.com/WORLD/9802/24/s.korea.wrap/index.html`.

Dafoe, Allan and Devin Caughey. 2016. "Honor and War: Southern U.S. Presidents and the Effects of Concern for Reputation." *World Politcs* 68(2):Forthcoming.

Downs, George W. and David M. Rocke. 1995. *Optimal Imperfection?* Princeton University Press.

Egesdal, Michael, Zhenyu Lai and Che-Lin Su. 2013. "Estimating Dynamic Discrete-Choice Games of Incomplete Information." Unpublished manuscript.

Fearon, James D. 1994*a*. "Domestic Political Audiences and the Escalation of International Disputes." *American Political Science Review* 88(3):577–592.

Fearon, James D. 1994*b*. "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. 2004. "Why Do Some Civil Wars Last So Much Longer than Others?" *Journal of Peace Research* 41(3):275–301.

Feenstra, Robert C., Robert Inklaar and Marcel P. Timmer. N.d. "The Next Generation of the Penn World Table." available for download at www.ggdc.net/pwt.

Gartzke, Erik. 1998. "Kant We All Just Get Along? Opportunity, Willingness, and the Origins of the Democratic Peace." *American Journal of Political Science* 42(1):1–27.

Gent, Stephen E. 2007. "Strange Bedfellows: The Strategic Dynamics of Major Power Military Interventions." *Journal of Politics* 69(4):1089–1102.

Ghosn, Faten, Glenn Palmer and Stuart Bremer. 2004. "The MID3 Data Set, 1993-2001: Procedures, Coding Rules, and Description." *Conflict Management and Peace Science* 21(2):133–154.

Gleditsch, Kristian S. 2002. "Expanded Trade and GDP Data." *Journal of Conflict Resolution* 46:712–724.

Gleditsch, Kristian S. and Michael D. Ward. 2000. "War and Peace in Space and Time: The Role of Democratization." *International Studies Quarterly* 44(1):1–29.

Hotz, V. Joseph and Robert A. Miller. 1993. "Conditional Choice Probabilities and the Estimation of Dynamic Models." *The Review of Economic Studies* 60(3):497–529.

Huth, Paul K. and Todd L. Allee. 2002. "Domestic Political Accountability and the Escalation and Settlement of International Disputes." *Journal of Conflict Resolution* 46(6):754–790.

Jo, Jinhee. 2011. "Nonuniqueness of the Equilibrium in Lewis and Schultz's Model." *Political Analysis* 19(3):351–362.

Karlekar, Karin Deutsch and Jennifer Dunham. 2012. "Freedom of the Press 2012." Freedom House. Published at: `http://www.freedomhouse.org/report-types/freedom-press`.

Kenwick, Michael R., Matthew Lane, Benjamin Ostick and Glenn Palmer. 2013. "Codebook for the Militarized Interstate Incident Data, version 4.0." Unpublished manuscript.

Kurizaki, Shuhei. 2007. "Efficient Secrecy: Public Versus Private Threats in Crisis Diplomacy." *American Political Science Review* 101(3):543–558.

Kurizaki, Shuhei and Taehee Whang. 2015. "Detecting Audience Costs in International Disputes." *International Organization* 69(4):949–980.

Lavelle, Moira. 2015. "A brief history of border conflict between North and South Korea." *PRI's The World*. Retrived from: `http://www.pri.org/stories/2015-08-20/brief-history-border-conflict-between-north-and-south-korea`.

Lee, Manwoo. 2002. "Sunset for Kim Dae-jung's Sunshine Policy?" *Current History* 101(654):161–171.

Lektzian, David and Mark Souva. 2003. "The Economic Peace Between Democracies: Economic Sanctions and Domestic Institutions." *Journal of Peace Research* 40(6):641–659.

Leventoglu, Bahar and Ahmer Tarar. 2005. "Prenegotiation Public Commitment in Domestic and International Bargaining." *American Political Science Review* 99(3):419–433.

Levy, Jack S. 2012. "Coercive Threats, Audience Costs, and Case Studies." *Security Studies* 21(3):383–390.

Levy, Jack S., Michael K. McCoy, Paul Poast and Geoffrey P.R. Wallace. 2015. "Backing out or Backing in? Commitment and Consistency in Audience Costs Theory." *American Journal of Political Science* 59(4):988–1001.

Lewis, Jeffrey B. and Kenneth A. Schultz. 2003. "Revealing Preferences: Empirical Estimation of a Crisis Bargaining Game with Incomplete Information." *Political Analysis* 11(4):345–367.

Li, Quan. 2005. "Does Democracy Promote or Reduce Transnational Terrorist Incidents?" *Journal of Conflict Resolution* 49(2):278–297.

Magnac, Thierry and David Thesmar. 2002. "Identifying Dynamic Discrete Decision Processes." *Econometrica* 70(2):801–816.

McKelvey, Richard D. and Thomas R. Palfrey. 1995. "Quantal Response Equilibria for Normal Form Games." *Games and Economic Behavior* 10(1):6–38.

Oneal, John R. and Bruce Russett. 1997. "The Classical Liberals Were Right: Democracy, Interdependence, and Conflict, 1950-1985." *International Studies Quarterly* 41(2):267–293.

Paik, Haksoon. 2002. "Assessment of the Sunshine Policy." *Asian Perspective* 26(3):13–35.

Pakes, Ariel, Michael Ostrovsky and Steven Berry. 2007. "Simple Estimators for the Parameters of Discrete Dynamic Games (with entry/exit examples)." *RAND Journal of Economics* 38(2):373–399.

Partell, Peter J. and Glenn Palmer. 1999. "Audience Costs and Interstate Crises: An Emprical Assessment of Fearon's Model of Dispute Outcomes." *International Studies Quarterly* 43(2):389–405.

Pesendorfer, Martin and Philipp Schmidt-Dengler. 2003. "Identification and Estimation of Dynamic Games." National Bureau of Economic Research, No. w9726.

Prins, Brandon C. 2003. "Institutional Instability and the Credibility of Audience Costs: Political Participation and Interastate Crisis Bargaining, 1816-1992." *Journal of Peace Research* 40(1):67–84.

Regan, Patrick M., Richard W. Frank and David H. Clark. 2009. "New Datasets on Political Institutions and Elections, 1972–2005." *Conflict Management and Peace Science* 26(3):286–304.

Rust, John. 1987. "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher." *Econometrica* 55(5):999–1033.

Schultz, Kenneth A. 1998. "Domestic Opposition and Signaling in International Crises." *American Political Science Review* 92(4):829–844.

Schultz, Kenneth A. 1999. "Do Democratic Institutions Constrain or Inform? Contrasting Two Institutional Perspectives on Democracy and War." *International Organization* 53(2):233–266.

Schultz, Kenneth A. 2001. "Looking for Audience Costs." *Journal of Conflict Resolution* 45(1):32–60.

Schultz, Kenneth A., Jeffrey B. Lewis and Cesar Zucco. 2012. "Coercive Diplomacy Data Base." [Data File and Codebook].

Signorino, Curtis S. 1999. "Strategic Interaction and the Statistical Analysis of International Conflict." *American Political Science Review* 93(2):279–297.

Silvey, S. D. 1959. "The Lagrangian Multiplier Test." *The Annals of Mathematical Statistics* 30(2):389–407.

Singer, J. David, Stuart A. Bremer and John Stuckey. 1972. Capability Distribution, Uncertainty, and Major Power War, 1820-1965. In *Peace, War, and Numbers*, ed. Bruce Russett. Beverly Hills: Sage Publications.

Slantchev, Branislav L. 2006. "Politicians, the Media, and Domestic Audience Costs." *International Studies Quarterly* 50(2):445–477.

Slantchev, Branislav L. 2012. "Audience Cost Theory and Its Audiences." *Security Studies* 21(3):376–382.

Smith, Alastair. 1998. "International Crises and Domestic Politics." *American Political Science Review* 92(3):623–638.

Snyder, Jack and Erica D. Borghard. 2011. "The Cost of Empty Threats: A Penny, Not a Pound." *American Political Science Review* 105(3):437–456.

Su, Che-Lin and Kenneth L. Judd. 2012. "Constrained Optimization Approaches to Estimation of Structural Models." *Econometrica* 80(5):2213–2230.

Tarar, Ahmer and Bahar Leventoğlu. 2009. "Public Commitment in Crisis Bargaining." *International Studies Quarterly* 53(3):817–839.

Thompson, William R. and David R. Dreyer. 2012. *Handbook of International Rivalries: 1494-2010*. Los Angeles: CQ Press.

Tomz, Michael. 2007. "Domestic Audience Costs in International Relations: An Experimental Approach." *International Organization* 61(4):821–840.

Uzonyi, Gary, Mark Souva and Sona N. Golder. 2012. "Domestic Institutions and Credible Signals." *International Studies Quarterly* 56(4):765–776.

Weeks, Jessica L. 2008. "Autocratic Audience Costs: Regime Type and Signaling Resolve." *International Organization* 62(1):35–64.

Weeks, Jessica L. 2012. "Strongmen and Straw Men: Authoritarian Regimes and the Initiation of International Conflict." *Amercian Political Science Review* 106(2):326–347.

Whang, Taehee, Elena V. McLean and Douglas W Kuberski. 2013. "Coercion, Information, and the Success of Sanction Threats." *American Journal of Political Science* 57(1):65–81.

Wolf, Albert B. 2016. "Backing Down and Domestic Political Survival in Israel." *Polity* 48(3):414–439.