Audience Costs and Reputation in Crisis Bargaining☆

Selçuk Özyurt[a]

[a]*Sabancı University, Faculty of Arts and Social Sciences, 34956, Orhanlı, Istanbul, Turkey*

Abstract

In crisis bargaining literature, it is conventional wisdom that the ability of generating higher audience costs is an advantage for a leader of a state. However, empirical studies show that democratic states use this mechanism only occasionally. This paper formally shows that higher audience costs may be good or bad depending on (1) the benefit-cost ratio of the crisis, (2) initial probability of resolve, and (3) how fast states generate audience costs with time. In particular, if the value of the prize over the cost of attacking is low or the initial probability of resolve is high enough, then having greater ability to generate audience costs may undermine democratic states' diplomatic success.

Keywords: Crisis bargaining, Audience costs, Reputation, Behavioral types

1. Introduction

A growing literature on crisis bargaining argues that the presence of domestic audiences is the major source of diplomatic success. The idea that some leaders have an easier time generating audience costs is advanced in the seminal paper of James Fearon (1994) as a plausible working hypothesis that has interesting theoretical and empirical implications. In a world with audience costs, the risk of losing public support or even office, signaling incompetence, or losing international/national credibility could discourage leaders from making empty threats and promises. That is, audience costs may occur if the leader makes public

☆This research was supported by the Marie Curie International Reintegration Grant (no. 256486) within the European Community Framework Programme. I would like to thank the advisory editor and two anonymous referees for their helpful comments and suggestions. All the remaining errors are my own.

Email address: `ozyurt@sabanciuniv.edu` (Selçuk Özyurt)

threats or commitments but fails to carry through on them.[1] They help because leaders that are more *sensitive to audience costs* are less likely to bluff, and thus, the threats they make are more likely to be genuine. As a result, the targets of their challenges should be less likely to resist.

It is conventional wisdom that the ability of generating higher audience costs is an advantage for a leader, and thus, making a firm public stand strengthens a country's position and gives it the upper hand in a crisis. This idea actually emerges from Fearon's original model of crisis bargaining. In that model, each state's (ex-ante) expected payoff increases with the state's sensitivity to audience costs. Fearon explicitly states this point in his paper in several occasions and adds (page 585), "This result provides a rationale for why, ex-ante, both democratic and authoritarian leaders would want to be able to generate significant audience costs in international contests."

However, despite the sincere efforts that have been made to support the conjectures derived from the audience costs theory, empirical studies could not help but increase the skeptic views about the validity of this theoretically plausible mechanism. The works of Trachtenberg (2012) and Snyder and Borghard (2011), two recent empirical papers, discuss that leaders make use of this mechanism infrequently. For example, in 1990, prior to the Gulf War, President Bush made explicit foreign policy statements regarding Iraq's invasion of Kuwait and threatened Saddam Hussein. By comparison, U.S. policy toward Bosnia was less direct. Both presidents Bush and Clinton adopted vague, ambiguous policies toward the Bosnian crisis, and U.S. intervention was limited (Smith 1998). Likewise, the United Kingdom made very mild threats to Iran during the Abadan Crisis of 1951 but issued serious warnings to Argentina during the Falkland Islands crisis in 1982.

In light of analogous historical examples, it could be the case—as opposed to the common belief—that audience costs do not help leaders in all crises. This paper formally proves this. In particular, I show that if the value of the good in dispute is low or if war is not expected to go well, then having greater sensitivity to audience costs could be bad.[2]

[1]For further discussions about the microfoundations of audience costs, see Smith (1998), Guisinger and Smith (2002), Schultz (2001), and Slantchev (2006). For empirical evidence that audience costs exist, see Eyerman and Hart (1996), Gelpi and Griesdorf (2001), Partell and Palmer (1999), Prins (2003), Leventoglu and Tarar (2005), and Tomz (2007).

[2]Schultz (2012) supports this argument with no formal treatment.

For this purpose, I use the crisis bargaining game of Fearon (1994) with no structural change, except the modeling of the asymmetric information. In the spirit of Harsanyi (1967), I assume that each state has two types: *resolved* and *flexible* (i.e., not resolved). A resolved leader is stubborn and committed not to back down. The flexible type has no commitment to his demand and is willing to back down and mimic the resolved type (i.e., bluff) if optimal. This interpretation of resolve is consistent with that of Schelling (1966): he suggests that states may be resolved or not and that they wish to build reputation in order to increase their credibility in international conflicts.

Readers who are not familiar with the technical details of Fearon's crisis bargaining game may prefer to skip the next subsection and directly go to Section 2, where I explain the details of the infinite-horizon, continuous-time game. Section 3 provides the main results of this paper. In Section 3.1, I compare and contrast Fearon's original results with the current one(s). Section 4 discusses some simple extensions, showing that we need to distinguish between the *level* of audience costs and the *derivative* of audience costs. In particular, I show in Section 4 that a higher level of audience costs is always good, but a higher derivative of audience costs is not. Finally, I conclude in Section 5.

1.1. Overview of the Model and of the Results

Fearon's crisis bargaining game is an infinite-horizon, continuous-time game between two states. These two players are in a dispute over a prize worth $v_i > 0$ for each, and at all times $t \geq 0$ before the crisis ends, each player can choose to escalate, back down (quit), or attack. The crisis ends when one or both states attack or quit. If either state attacks before the other quits, the dispute ends with war, and each state i receives the (net expected) payoff of $-w_i < 0$, indicating all the risks and gains in a military intervention. If state i quits at time $t \geq 0$ before the other has backed down or attacked, then its adversary j receives the prize, while state i suffers audience costs equal to $c_i(t)$, a continuous and strictly increasing function of the amount of escalation t. Fearon supposes for simplicity that each state's audience costs function is linear on time, and so $c_i(t) = c_i t$ with $c_i > 0$. Therefore, state 1 is more sensitive to audience costs if and only if $c_1 > c_2$.

Fearon (1994) assumes that each state i knows its own cost of war (i.e., w_i) but knows only the distribution of its adversary's cost. Both players are willing to quit in Fearon's model if doing so is optimal for them. That is, Fearon does not assume resolved types (who never quit the crisis bargaining game) a priori. However, in equilibrium, a player that has low cost of war prefers to play a strategy in which he never quits. Therefore, the stubborn behavior—or the resolved type—is endogenously implied by the equilibrium in Fearon (1994). On the other hand, the model in this paper supposes that each state holds a positive prior belief that its rival is resolved. Once the dispute becomes public (i.e., after time 0), these initial priors are updated according to the flexible players' equilibrium strategies, which depend on the prospects of the crisis (such as the players' value for the prize and the cost of war). That is, flexible players build reputation on their resolve in order to increase their credibility in the crisis bargaining game.

A leader who is engaged in a dispute with another state may be resolved because of his firm belief that backing down, and thus giving up for the prize, is simply a decision that will not be ratified by domestic actors (Iida 1993). Or the leader may want to signal competence and increase his/her chance of reelection. Therefore, a leader may commit not to back down in a crisis for reasons other than the prospects of the crisis, and positive priors about states' resolve may be interpreted as the states' initial beliefs on the existence of such motives. Unlike Fearon (1994), this alternative modeling implies cases where a state is resolved even when the prospects of the crisis are low and a state is flexible even if the prospects of the crisis are high.

The main implication of this modeling is that Fearon's conclusion—higher sensitivity to audience costs is good—may not hold when the players' prior beliefs, relative to the prospects of the crisis, are high enough. In the current model, there are two commitment devices that flexible players can use to make their bluff credible. The first one is the cost of war and the ability of generating audience costs. The second one is the players' reputation on resolve. In equilibrium, both states play a mixed strategy, but the state that is less sensitive to audience costs (call it state 2) quits with a greater rate. Therefore, state 2 builds its reputation faster.[3] Building reputation requires time, and so, reputation for resolve is a credible commitment device when states expect a long horizon for the crisis or have high initial reputation. Therefore, if the relative likelihood of a state being the resolved type exceeds the benefit-cost ratio of the crisis, then the states are no longer in the Fearon's regime and so state 2 is advantaged. This case never occurs in Fearon's model because his

[3]Given its equilibrium strategy, a player's reputation at time t is the conditional probability that this player is the resolved type, conditional on the event that the player has not yet quit prior to time t.

modeling of the uncertainty (or resolve) indirectly gives the reputational advantage to the state that is more sensitive to audience costs.[4]

2. The Model

Two states (or leaders), 1 and 2, are in dispute over a prize (e.g., territory, valuable resources, or austerity measures) worth $v_i > 0$ for each state $i \in \{1, 2\}$. The crisis occurs in continuous time. At all times $t \geq 0$ before the crisis ends, each state can choose to escalate (wait), back down (quit), or attack. The crisis ends when one or both states attack or quit.

Payoffs are given as follows: If either state attacks before the other quits, the dispute ends with war, and each state i receives the (net expected) payoff of $-w_i < 0$. It indicates all the risks and gains in a military intervention. That is, w_i is the cost of attacking less the expected benefit of attacking.[5] The inequality $-w_i < 0$ indicates that not being involved in the dispute is more desirable than attacking. If state i concedes at time t before the other has backed down or attacked, then its opponent j receives the prize, while state i suffers audience costs equal to $c_i(t)$, a continuous and strictly increasing function of the amount of escalation with $c_i(t) \geq 0$ for all t. Only the state that backs down experiences audience costs.

If one state chooses to attack at time t and the other chooses to quit or attack at the same time, both states receive their war payoffs, $-w_i$. However, if both quit at time t, then state i receives $\frac{v_i}{2} - c_i(t)$. These particular assumptions are not crucial because in equilibrium, simultaneous concessions or attacks occur with probability 0. Finally, if states escalate the conflict indefinitely, each state gets a payoff that is strictly less than its war payoff. Players' payoff at time infinity would be interpreted as payoff of "perpetual conflict avoidance." By neglecting to address high-conflict situations, avoiders risk allowing problems to fester out of their control. By assuming that players' payoffs of perpetual conflict avoidance are worse than their war payoff, we ensure that the leaders are confrontational. That is, the leaders prefer confrontation even if the conflict may end badly for them.

Finally, I assume that there is some uncertainty about states' *resolve*. In the spirit of

[4] See Section 3.1 for a detailed discussion.

[5] Note that this sign convention for w_i is the reverse of the one that is used by Fearon (1994).

Harsanyi (1967), each state has two types: *resolved* or *flexible*. The prior probability that state i is resolved is denoted by $z_i \in (0, 1)$. Each state knows its own type but does not know the opponent's true type. The flexible types pick a strategy, given their beliefs, to maximize their expected payoff in the crisis bargaining game. On the other hand, a resolved type never backs down and never attacks. That is, a resolved type always chooses the action "escalate."[6] Call this *crisis game* where all the parameters are common knowledge G.

A leader may commit not to back down in a crisis for reasons other than the prospects of the crisis. For example, a leader may be resolved because of his firm belief that backing down, and thus giving up for the prize, is simply a decision that will not be ratified by domestic actors (Iida 1993). Therefore, the positive priors, z_i's, can be interpreted as the players' initial beliefs on the existence of such motives that may force the leaders not to quit.

Resolved types closely resemble the obstinate types that are first defined by Myerson (1991) (r-insisting types) and studied further by Abreu and Gul (2000) and Kambe (1999).[7] In a bilateral negotiation, obstinate types always demand a particular share and accept an offer if and only if it weakly exceeds that share. Abreu and Sethi (2003) support the existence of obstinate types from an evolutionary perspective and show that if players incur a cost of rationality, even if it is very small, the absence of such types is not compatible with evolutionary stability in a bargaining environment.

In the context of international crises, audience costs can be interpreted as the *opportunity cost* of backing down. That is, it is the total value of the opportunities that are missed or not used effectively by backing down from the hard-line policy. In the crisis bargaining literature, there is a considerable amount of discussion concerning microfoundations of audience costs. For instance, a leader that backs down from a public commitment may suffer audience costs because the leader is less likely to be reelected (Smith 1998), has violated national honor (Fearon 1994), or has lost his reputation and the credibility of his rhetoric (Sartori 2002, Guisinger and Smith 2002).

[6] One can suppose that the resolved types never quit and attack only at time $K \in \mathbb{R}_+$. Clearly, if $K = 0$ or very close to 0, for example, then it will be a binding bound and thus alter flexible players' equilibrium strategies. However, if K is sufficiently large—in particular, larger than t^*, which is defined in Section 3—the results of the paper will not be affected.

[7] These "strategy types" are first used in game theory by Kreps and Wilson (1982) and Milgrom and Roberts (1982), where commitments are modeled as behavioral types that exist in society so that rational players can mimic them if it is optimal to do so.

This interpretation provides additional insight into Fearon's critical assumption that audience costs increase with time. Public statements or messages that the leaders send in crises are expected to serve an agenda states form. This agenda may include items that are not directly related to the prize, such as increasing audiences' support and the likelihood of winning elections, building reputation for future negotiations, discouraging potential rivals, and preventing future crises. Therefore, a leader's expected value of following a hard-line policy may increase with time because as time passes, the leader is more likely to convince domestic and international audiences about his resolve, and persuaded audiences would increase the likelihood of the successful execution of his agenda. As a result, if the expected benefit of following the hard-line policy increases in time, then the opportunity cost of backing down (i.e., audience costs) is expected to increase with time as well.

2.1. Strategies

The strategies of the resolved types are simple: never back down and never attack. On the other hand, flexible types have the opportunity of credibly bluffing, manipulating the adversaries' belief, and thus building reputation as a leverage to increase their bargaining power. Therefore, what motivates and interests the subsequent analyses are the optimal (equilibrium) strategies of the flexible types.

A strategy $\sigma_i = (Q_i, A_i)$ of state i is defined by a pair of cumulative distributions. $Q_i(t) : \mathbb{R}_+ \cup \{\infty\} \to [0,1]$ represents the probability that state i quits by time t (inclusive), and $A_i(t) : \mathbb{R}_+ \cup \{\infty\} \to [0,1]$ denotes the probability that state i attacks by time t (inclusive). If state i's strategy is $Q_i(t)$, then flexible state i's strategy is $Q_i(t)/(1-z)$ because the resolved types never quit. The same arguments hold for $A_i(t)$. Note that $Q_i(t) + A_i(t) \leq 1 - z_i$ for all $t \geq 0$.

Given Q_j and A_j, flexible state i's expected payoff of quitting at time t is

$$\begin{aligned} U_i(t, Q_j, A_j) := {} & v_i Q_j(t) - w_i A_j(t) + [1 - Q_j(t) - A_j(t)]\,[-c_i(t)] \\ & + \left(\frac{v_i}{2} - c_i(t)\right) \left[Q_j(t) - Q_j(t^-)\right], \end{aligned} \tag{1}$$

with $Q_j(t^-) = \lim_{y \uparrow t} Q_j(y)$.

3. Main Results

In this section, I analyze two special cases. These particular cases both convey the flavor of the analysis and are the basic building blocks for the extended versions of the model examined subsequently. For the rest of this section, I will assume that states' audience cost functions are linear $c_i(t) = tc_i$ and $c_1 > c_2 > 0$. Therefore, state 1 is more sensitive to audience costs. However, states are identical in all other aspects. That is, $v_i = v > 0$, $w_i = w > 0$, and $z_i = z$ for each i.

For the first special case, suppose that both states are known to be flexible (i.e., $z = 0$). This is also a special case of Fearon's model, and higher sensitivity to audience costs is always an advantage for diplomatic success. This conclusion follows from the fact that in the unique subgame perfect equilibrium of this special case, state 2 quits at time 0, state 1 never quits, and thus, payoffs of states 1 and 2 are v and 0, respectively.

In equilibrium, state i does not back down beyond time t_i, where i's audience costs are equal to its cost of war (i.e., $t_i c_i = w$). Since state 1 can generate higher audience costs (as $c_1 > c_2$), we have $t_1 < t_2$. That is, states never attack before t_1 and state 1 never quits after t_1. Therefore, if the conflict ever reaches time t_1, then subgame perfection implies that state 2 must quit immediately. Thus, the crisis bargaining game ends prior to time t_1. Moreover, there exists no equilibrium in which state 1 quits prior to time t_1. Suppose for a contradiction that there is an equilibrium in which state 1 quits at time t where $t \in [0, t_1]$. On the one hand, state 2's equilibrium strategy cannot dictate it to quit before time t because state 2 would achieve a higher payoff if it waited until time t and quitted at time $t+\epsilon$, where $\epsilon \geq 0$ is small. On the other hand, state 2's equilibrium strategy cannot dictate it to quit after time t (and before time t_1) because state 1 would achieve a higher payoff if state 1 waited until time t_1 and attacked afterwards. Therefore, in equilibrium, state 2 quits at time 0 because it anticipates that state 1 never quits and that delaying the concession until time t_1 has no benefit.

The second case is more interesting. Now I resume the case where z is positive (i.e., $z \in (0, 1)$). In this case, each flexible state has the option of building reputation on its resolve by bluffing and escalating the crisis. The equilibrium concept that I will use is sequential equilibrium. A notable implication of the equilibrium is that sensitivity to audience costs is not always an advantage for diplomatic success.

In equilibrium, all sequential equilibria have the same Q_i functions, and these functions are nondegenerate cumulative distribution functions (i.e., truly mixed quitting strategies).[8] In particular, each flexible state chooses the time of backing down randomly with a decreasing hazard rate. That is, flexible states are more likely to quit at the early stages of the crisis and less likely to quit as it escalates further. Moreover, escalation of the conflict stops at some finite (deterministic) time t^*, a function of the primitives, with certainty, and no state attacks before this time. Therefore, t^*—the equilibrium *horizon* of the game—is *the time beyond which no type quits*. In what follows, I will prove all these claims and the main results of the paper.

We describe state i's behavior in the crisis game by a probability distribution over quitting times, and we allow $Q_i(0) > 0$, so i may quit immediately with positive probability. Let $\lambda_i(t)$ be state i's instantaneous quitting (or hazard) rate at time t with the condition that no state has backed down or attacked before this time. Therefore, by definition, $\lambda_i(t) = \frac{dQ_i(t)/dt}{1-Q_i(t)}$. Since there is no pure strategy equilibrium,[9] we look for an equilibrium where flexible player j mixes between quitting and escalating. Flexible j is indifferent between quitting at time t and waiting for an infinitesimal period Δ and then quit at time $t+\Delta$ if and only if

$$-tc_j = v\lambda_i(t)\Delta - [1 - \lambda_i(t)\Delta](t+\Delta)c_j,$$

where $\lambda_i(t)\Delta$ is the probability that i quits in the interval Δ.[10] Solving this equation for $\lambda_i(t)$ and taking its limit as Δ approaches zero yields

$$\lambda_i(t) = \frac{c_j}{v + tc_j}.$$

Note that the last expression and the definition for $\lambda_i(t)$ imply the differential equation

$$-\frac{d}{dt}\ln(1 - Q_i(t)) = \frac{d}{dt}\ln(v + tc_j)),$$

which can be directly solved to yield

$$Q_i(t) = 1 - \frac{\Big(1 - Q_i(0)\Big)v}{v + tc_j}.$$

[8] There are still multiple sequential equilibria as the A_i's can vary.

[9] Proposition 1 proves that there is no pure strategy equilibrium.

[10] I assume, without loss of generality, that $c_i(t+\Delta) < w$ for $i = 1, 2$, so that both states prefer backing down over attacking before time $t + \Delta$.

Proposition 1. *The crisis game G has a sequential equilibrium. Any* $\sigma = \Big((Q_1, A_1), (Q_2, A_2)\Big)$ *is a sequential equilibrium of G if and only if $t^* \leq \frac{w}{c_1}$ exists such that, for $i = 1, 2$, the following conditions hold:*

1. $Q_i(t) = 1 - \frac{\big(1-Q_i(0)\big)v}{v+tc_j}$ *for all* $t \leq t^*$,
2. $Q_1(0)Q_2(0) = 0$,
3. $Q_i(t^*) = 1 - z$ *if* $t^* < w/c_1$ *and* $Q_1(t^*) \leq Q_2(t^*) = 1 - z$ *if* $t^* = w/c_1$, *and*
4. $A_i(t) = 0$ *for all* $t < t^*$ *and* $\lim_{k\to\infty} A_i(k) = 1 - z - Q_i(t^*)$.

Proof. Let $\sigma = \Big((Q_1, A_1), (Q_2, A_2)\Big)$ define a sequential equilibrium. I will argue that σ must have the form specified and that these strategies do indeed define an equilibrium. For $i = 1, 2$, let $\kappa_i = \inf\{t \geq 0 | Q_i(t) = \lim_{k\to\infty} Q_i(k)\}$, where $\inf \emptyset := \infty$. That is, κ_i denotes the time beyond which state i does not quit. Then:[11]

(*i*) $\kappa_1 = \kappa_2$: A flexible player does not delay quitting once he knows that his opponent will never quit.

(*ii*) If Q_i jumps at t, then Q_j is constant at some ϵ-neighbourhood of t: the reason is that j would always want to wait for some small $\epsilon > 0$ amount of time in order to enjoy the discrete chance of i quitting.

(*iii*) If Q_i is constant between (t', t''), then so is Q_j: if i will not quit between (t', t''), then if j plans to quit in this interval, he does better to quit immediately at t' rather than wait to some time $t > t'$.

(*iv*) There is no interval (t', t'') with $t'' < \kappa_1$ on which Q_1 and Q_2 are constant: if so, i would do better to quit at $t'' - \epsilon$ than to quit at t'', leading to contradiction.

(*v*) If $t' < t'' < \kappa_1$, then $Q_i(t'') > Q_i(t')$ for $i = 1, 2$: as we noted in (*iii*), if Q_i is constant on some interval (t', t''), then the optimality of Q_j implies that Q_j is also constant on (t', t''). Therefore, the conditions (*iii*) and (*iv*) imply (*v*).

[11] Proofs of the arguments (*i*) $-$ (*v*) directly follow from the proof of Proposition 1 in Abreu and Gul (2000) and are analogous to the arguments in Hendricks, Weiss and Wilson (1988). Therefore, I skip the details.

From $(i)-(v)$, it follows that Q_1 and Q_2 must be continuous and strictly increasing on $[0,t^*]$, where $t^*=\kappa_1=\kappa_2$. That is, flexible players are indifferent between quitting at time $t<t^*$ and waiting for an infinitesimal period Δ and then quit at time $t+\Delta<t^*$. But then, players must be quitting with the hazard rates λ_1,λ_2 that are defined above. Thus, as derived previously, $Q_i(t)=1-\frac{(1-Q_i(0))v}{v+tc_j}$ for all $t\leq t^*$. By (ii), both $Q_1(0)$ and $Q_2(0)$ cannot be positive, implying that $Q_1(0)Q_2(0)=0$.

Since Q_i is strictly increasing on $[0,t^*]$, it must be the case that $t^*\leq w/c_1<w/c_2$. Therefore, $A_i(t)=0$ for all $t<t^*$. Since escalating forever is costlier than attacking and $Q_i(t)+A_i(t)\leq 1-z$ for all $t\geq 0$, we must have $\lim_{k\to\infty}A_i(k)=1-z-Q_i(t^*)$ for $i=1,2$. Finally, since a flexible player will not delay quitting once he knows his opponent will never quit and will not attack before time t^*, we can conclude that $Q_i(t^*)=1-z$ for $i=1,2$ if $t^*<w/c_1$. However, when $t^*=w/c_1$, state 1 (and only state 1) will be indifferent between quitting at time t^* and attacking at (or after) time t^*. Therefore, we must have $Q_1(t^*)\leq Q_2(t^*)=1-z$ if $t^*=w/c_1$.

Finally, suppose now that the strategies $\sigma=\Big((Q_1,A_1),(Q_2,A_2)\Big)$ satisfy the conditions that are given by **1-4**. Note that the value of t^* is determined by the conditions **1-3** (see Lemma 1). Recall the derivation of the hazard rates λ_1 and λ_2. Given Q_j, flexible state i is indifferent between quitting at time t' and waiting for some time and then quitting at time t'', where $0\leq t'<t''\leq t^*$. Hence, any mixed strategy on the support $[0,t^*]$, in particular, Q_i is optimal for player i. According to the strategies, $Q_i(t^*)=1-z$. This is optimal for flexible state i when $t^*<w/c_1<w/c_2$. In addition, if $t^*=w/c_1$, then state 1 will be indifferent between quitting at time t^* and attacking at some time after (or at) time t^*. Therefore, $Q_1(t^*)\leq 1-z$ is optimal for flexible state 1 if $t^*=w/c_1$. According to the σ, state 2 never attacks (i.e., $A_2(t)=0$ for all t), whereas state 1 may attack with a positive probability at some time after t^*. Since the timing of attacking beyond t^* does not change flexible state 1's payoff, any A_1 satisfying condition **4** is also optimal. Hence, σ is indeed an equilibrium.

□

State i's hazard rate $\lambda_i(t)=c_j/(v+tc_j)$ is decreasing with time. Since j's audience costs increases with time, i's quitting (hazard) rate must be bigger at earlier times of the escalation to make j indifferent between quitting and escalating at all times. Note that the hazard rate depends upon only two parameters: (1) the value of the prize and (2) the opponent's audience costs coefficient. Therefore, *higher values for the prize will make* *states' concession rates lower*, and this prediction is consistent with intuition. Moreover, since $c_1 > c_2$, we have $\lambda_1(t) < \lambda_2(t)$. That is, state 2 quits with a greater rate. Therefore, conditional on that $Q_1(0) = Q_2(0) = 0$ and that the game G reaches time $t > 0$, the posterior probability that state 2 is resolved will be greater than the posterior probability that state 1 is resolved. The last observation implies that state 2 can build its reputation much faster if $Q_i(0)$s are the same, and thus, it is easier for state 2 to convince the adversary (about its resolve) if it bluffs. As I will show in the next proposition, this reputational advantage of the second state will put it into an advantageous position[12] even though state 2 is in an unfavorable condition regarding its ability to generate audience costs.

The following two lemmata solve the equilibrium horizon of the crisis game t^* as a function of the primitives and provide the unique equilibrium strategies $Q_1(t)$ and $Q_2(t)$. The equilibrium value of t^* can be derived by conditions **1-3** in Proposition 1. However, instead of providing this mechanical derivation, I will present an alternative proof that elucidates the equilibrium dynamics of the crisis bargaining game G, in particular, players' reputational concerns.

Lemma 1. *In a sequential equilibrium, the crisis bargaining game G ends by time* $t^* = \min\{t_1^*, t_2^*\}$, *where* $t_i^* = \min\left\{\frac{w}{c_i}, \frac{v(1-z)}{zc_j}\right\}$ *for* $i, j \in \{1, 2\}$ *with* $j \neq i$.

Proof. In equilibrium, if state i believes that j will never quit after time t and $c_i(t) < w$, then the flexible i will immediately quit at this time. There are two critical thresholds beyond which i believes that state j will never quit. One of them is the time that state j's reputation reaches 1, and the other is the time t satisfying $tc_j = w$.

Given that j does not quit at time 0 with a positive probability (i.e., $Q_j(0) = 0$), state i will be convinced that j is resolved by time τ_j, solving $Q_j(\tau_j) = 1 - \frac{v}{v+\tau_j c_i} = 1 - z$, implying $\tau_j = \frac{v(1-z)}{zc_i}$. In equilibrium, $Q_j(0)$ would take values more than 0. Therefore, i will be convinced that j is resolved no later than τ_j. Hence, state i never backs down after time $t_j^* = min\left\{\frac{w}{c_j}, \frac{v(1-z)}{zc_i}\right\}$. Similar arguments hold for state j, that is, flexible i never quits after time $t_i^* = min\left\{\frac{w}{c_i}, \frac{v(1-z)}{zc_j}\right\}$.

When, for example, $t_1^* < t_2^*$ and the conflict escalates until time t_1^*, flexible 2 ends the game at this time for sure because he knows that state 1 will never quit beyond this point.

[12] This term will be formally defined later.

As a result, the crisis bargaining game will end no later than $t^* = \min\{t_1^*, t_2^*\}$. Next, I will show that the game G cannot end before time t^*. Suppose that the game ends at time $\hat{t}$, where $\hat{t} < t^*$. According to the equilibrium strategies given by Proposition 1, $Q_2(0) > 0$ must hold. Otherwise, state 2's reputation will not reach 1 at time $\hat{t}$ (i.e., $Q_2(\hat{t}) = 1 - z$ will not hold), which contradicts the optimality of Q_2. With the same reasoning, we must have $Q_1(0) > 0$ because otherwise $Q_1(\hat{t}) = 1 - z$ will not hold (note that $\hat{t} < t_1^* \leq w/c_1$). However, having $Q_1(0) > 0$ and $Q_2(0) > 0$ simultaneously contradicts the optimality of Q_is (recall condition 2 of Proposition 1.) Hence, in equilibrium, given that flexible players randomize the timing of quitting, escalation continues until time $t^* = \min\{t_1^*, t_2^*\}$ with some positive probability and stops at this time with certainty.

□

Lemma 2. *In a sequential equilibrium, if $t_i^* > t_j^*$, then*

$$Q_j(t) = 1 - \frac{v}{v + tc_i} \quad \text{and} \quad Q_i(t) = 1 - \frac{(1 - Q_i(0))v}{v + tc_j},$$

where

$$Q_i(0) = \begin{cases} (1-z)[1 - \frac{zc_j}{c_i}], & \text{if } \frac{w}{c_j} > \frac{v(1-z)}{zc_i} \\ 1 - z - \frac{zw}{v}, & \text{otherwise.} \end{cases}$$

Proof. Since $t_i^* > t_j^*$, we have $t^* = t_j^*$. Therefore, $\frac{v(1-z)}{zc_j} = \tau_i > t^*$ and $w/c_i > t^*$. Recall that τ_i is the time satisfying $Q_i(\tau_i) = 1 - z$ if $Q_i(0) = 0$. Since the game ends before time τ_i and $w/c_i > t^*$ holds, we must have $Q_i(t^*) = 1 - z$. Hence, we have $Q_i(0) > 0$. According to the second condition of Proposition 1, we must have $Q_1(0)Q_2(0) = 0$. The last condition with $Q_i(0) > 0$ implies that we must have $Q_j(0) = 0$.

$Q_j(0) = 0$ implies $Q_j(t) = 1 - \frac{v}{v+tc_i}$. Moreover, since $Q_i(t_j^*) = 1 - \frac{(1-Q_i(0))v}{v+t_j^* c_j} = 1 - z$, we have $Q_i(0) = 1 - z - \frac{c_j t_j^* z}{v}$, where $t_j^* = \min\left\{\frac{w}{c_j}, \frac{v(1-z)}{zc_i}\right\}$. This completes the proof.

□

Note that if $t_i^* > t_j^*$, then in equilibrium, state i (and only state i) quits with a positive probability at time 0. That is, $Q_i(0) > 0$ and $Q_j(0) = 0$.

Definition 1. *State i is called **advantaged** in the crisis bargaining game G if flexible i's expected payoff is strictly positive in any sequential equilibrium of the game G.*

In equilibrium with horizon t^*, states' strategies entail quitting at any time before t^*. Using Equation (1) and Proposition 1, we see that for any small $\epsilon > 0$, the equilibrium payoff of flexible state i is

$$\begin{aligned} U_i &= vQ_j(\epsilon) - wA_j(\epsilon) + [-c_i(\epsilon)]\,[1 - Q_j(\epsilon) - A_j(\epsilon)] + \left[\frac{v}{2} - c_i(\epsilon)\right][Q_j(\epsilon) - Q_j(\epsilon^-)] \\ &= vQ_j(\epsilon) + [-c_i(\epsilon)][1 - Q_j(\epsilon)]. \end{aligned}$$

Taking $\epsilon \to 0$ thus yields

$$\begin{aligned} U_i &= vQ_j(0) + [-c_i(0)][1 - Q_j(0)] \\ &= vQ_j(0), \end{aligned} \tag{2}$$

We know from Proposition 1 that $Q_1(0)Q_2(0) = 0$ must hold in equilibrium. Thus, for all parameter values, at most one state is advantaged in the crisis bargaining game G. Suppose for now that the parameters of the game G satisfy $t_i^* > t_j^*$. Therefore, according to the equilibrium strategies given in Lemma 2, $U_i = 0$, whereas $U_j = vQ_i(0) > 0$. Thus, the following result immediately follows from these discussions.

Corollary 1. *State j is advantaged in the crisis bargaining game G if and only if the parameters of the game satisfy $t_i^* > t_j^*$.*

Proposition 2. *State 1—the one that is more sensitive to audience costs—is advantaged in the crisis bargaining game G if and only if the inequality*

$$\frac{v}{w} > \frac{z}{1-z}$$

holds, i.e., if and only if the benefit-cost ratio, v/w, exceeds the relative likelihood, $z/(1-z)$, of a state being the resolved type.

Proof. As we previously discussed, state 1 is advantaged in the game G if and only if $t_1^* < t_2^*$. Since $c_1 > c_2$, we have $\frac{w}{c_1} < \frac{w}{c_2}$, and $\tau_1 = \frac{v(1-z)}{zc_2} > \tau_2 = \frac{v(1-z)}{zc_1}$. Thus, $t_1^* < t_2^*$ holds if and only if $\frac{w}{c_1} < \frac{v(1-z)}{zc_1}$. The last inequality yields the desired result. With a similar reasoning, if $\frac{w}{c_1} > \frac{v(1-z)}{zc_1}$ holds, then we have $t_2^* < t_1^*$, implying that state 2 is advantaged. $\square$

Remark 2.. Note that state 2—the one that is less sensitive to audience costs—is advantaged in the crisis bargaining game G whenever the benefit-cost ratio is smaller than the relative likelihood of a state being the resolved type (i.e., $\frac{v}{w} < \frac{z}{1-z}$).

Being the advantaged state does not mean that this state will get the prize with certainty. In equilibrium, both states can get the prize or suffer audience costs with positive probabilities. However, for any $t \leq t^*$, the probability that the advantaged state gets the prize prior to time t is higher. Proposition 2 implies, in contrast with conventional wisdom, that *the ability of generating greater audience costs is not always an advantage.* Depending on the parameter values, equilibrium can be grouped into two categories: The first one, which includes all the values of the parameters satisfying $\frac{v}{w} > \frac{z}{1-z}$, is a regime such that state 1 is advantaged. In this case, the horizon of the conflict is w/c_1. The second category, where the set of parameters satisfy $\frac{v}{w} < \frac{z}{1-z}$, is the second regime, where state 2 is advantaged. In this case, the horizon of the conflict is $\frac{v(1-z)}{zc_1}$. Note that when z is small, we are in the first regime, and things are as in Fearon. In particular, there is no discontinuity when resolved types are introduced into Fearon's model.

State 1's sensitivity to audience costs (in particular, c_1) negatively affects equilibrium horizon, and this is true regardless of the regime. Therefore, when states can generate greater audience costs, fewer escalatory steps are needed to credibly communicate one's preferences. Thus, crisis between democratic states should see significantly fewer escalatory steps than crisis between authoritarian states (Fearon 1994).

The probability of a peaceful initial resolution is the sum of the states' initial concessions (i.e., $Q_1(0) + Q_2(0)$). In regime 1, only the second state makes the initial concession. Therefore, $Q_1(0) = 0$ and $Q_2(0) = 1 - z - \frac{zw}{v}$ by Lemma 2. As a result, a higher benefit-cost ratio of the crisis—given that the regime does not change—increases the probability of a peaceful initial resolution. That is, *disputes with low cost of war or a high value for the prize are more likely to settle without any escalation. And if these disputes turn into public crises, then they are more likely to have less escalatory steps.*

The probability that flexible state 1 will initiate war in equilibrium is

$$q_A = \max\left\{1 - \frac{wc_2}{(1-z)[vc_1 + wc_2]},\ 0\right\}.$$

Recall that flexible states never initiate war in regime 2. Therefore, the probability of war (weakly) increases as the value of the prize increases. As the value of the prize is larger, the autocratic state—the one that is less sensitive to audience costs—is less likely to dispute the democratic state's threat. But if the crisis escalates, then the probability of war will be greater. Moreover, the probability of war decreases as the cost of war increases.

3.1. Comparison with Fearon's Model

The timing of the crisis bargaining game G is identical to Fearon's original game. The only difference between Fearon's and the current setup is the way they model uncertainty. Note that both models yield the same results for small values of z, in particular, when $\frac{v}{w} > \frac{z}{1-z}$. By contrast, Fearon's analysis does not hold, and the setup in this paper yields different results when it puts enough weight on the prospect of facing a resolved type that will never back down (a high enough z).

Fearon (1994) assumes that each state is uncertain about the opponent's cost of war but knows its own. That is, states' beliefs are represented by a distribution function F over the set of types $[0, \bar{w}]$, where $0 < \bar{w}$. Fearon shows, in Lemma 2, that in equilibrium with horizon t^*, the set of types is divided into three classes:

(*i*) All the types $w_i \in [0, \bar{w}]$ satisfying $w_i < c_i(t^*)$ never quit.

(*ii*) All the types $w_i \in [0, \bar{w}]$ satisfying $w_i > c_i(t^*)$ back down before time t^* with certainty.

(*iii*) The type $w_i \in [0, \bar{w}]$ satisfying $w_i = c_i(t^*)$ will be indifferent between quitting and attacking at time t^*.

Therefore, in equilibrium, audience costs *separate* the players (types) according to their willingness to quit. In equilibrium with horizon $t^* > 0$, the probability that state 1 follows a strategy in which he never quits is equal to $F(c_1(t^*))$ (i.e., the probability that $w_i \leq c_i(t^*)$). Likewise, the probability that state 2 follows a strategy in which he never quits is equal to $F(c_2(t^*))$. Since $c_1(t^*) > c_2(t^*)$ (at least for the linear audience costs case), then $F(c_1(t^*)) > F(c_2(t^*))$. *That is, the probability of facing a resolved type that will never back down is always greater for player 2. This makes state 1 advantaged regardless of the primitives.*

To be able to draw analogies between Fearon's model and the current one, consider an equilibrium with horizon $t^* > 0$. When z is low (in particular $\frac{v}{w} > \frac{z}{1-z}$ holds), then my model will generate dynamics that are similar to Fearon's original model. In equilibrium, we will have $w = c_1(t^*) > c_2(t^*)$, where $t^* = w/c_1$ since i's audience costs is $c_i(t) = c_i t$, and flexible state 1 will quit before t^* with some probability that is strictly less than 1. We know these from the analyses that are presented earlier in this section and are neither implied by nor derived from Fearon's analyses. Given these observations, one can use Fearon's above arguments and conclude that in equilibrium, the initial probability that state 2 is facing a resolved type is equal to $z + [1 - \frac{Q_1(t^*)}{(1-z)}]$, where $1 - Q_1(t^*)/(1-z) > 0$ denotes the probability that flexible state 1 does not quit. However, the probability that state 1 is facing a resolved type is exactly equal to z. Similar to Fearon (1994), the probability of facing a resolved type is greater for player 2, making state 1 advantaged.

On the other hand, if z is high (or $\frac{v}{w} < \frac{z}{1-z}$ holds), then we will have $w > c_1(t^*) > c_2(t^*)$, where $t^* = v(1-z)/zc_1$. In this case, flexible types of both states will certainly quit before time t^*. Once again, these observations are implied by the previous analyses of this section, not by Fearon's model. Given these outcomes and Fearon's above arguments, we can conclude that, in equilibrium, the initial probability of facing a resolved type is equal to z for both states. Note that z is the probability that a state is resolved, and thus, it is the lower boundary for the probability of facing a resolved type. Since both states face the resolved type with the same probabilities, state 1 will lose its favorable position that is implied in the previous case. Recall that state 2 can build its reputation at a faster rate since $\lambda_1(t) < \lambda_2(t)$ for all t, making state 2 advantaged. This case never occurs in Fearon's model because the probability of facing a resolved type is endogenously derived in equilibrium and the probability that player 2 faces a resolved type is always greater than the probability that player 1 faces a resolved type.

4. Level vs. Derivative of Audience Costs

The main result of this paper—that higher audience costs are not always good—may raise serious queries about the validity of the audience cost mechanism. The final arguments of the previous section technically explain why the results of the current paper and Fearon (1994) diverge. *The main result of this paper should not be interpreted against the audience costs mechanism.* On the contrary, the analyses indicate that we must distinguish between the ***level*** *of audience costs* (i.e., $c_i(t)$) and the ***derivative*** *of audience costs* (i.e., $c_i'(t)$).[13]

[13] I am grateful to the referees whose comments shaped the discussion in this section greatly.

Thus, generating higher audience costs may be good or bad depending on how states produce these costs. If a state can generate greater audience costs by keeping its derivative low, then this is good. However, generating audience costs simply by producing greater $c_i'(t)$ would be bad. In this section, I will slightly extend the model to clarify the distinct impacts of the level and the derivative of audience costs.

In this section, I let $c_i(t)$ to be a positive, increasing, and differentiable function of time. Similar to the analyses in the previous section, let $\lambda_i(t)$ be state i's quitting rate at time t. Then flexible j is indifferent between quitting at time t and $t+\Delta$ if and only if

$$-c_j(t) = v\lambda_i(t)\Delta - [1-\lambda_i(t)\Delta]c_j(t+\Delta).$$

Solving this equation for $\lambda_i(t)$ and taking its limit as Δ approaches 0 yields

$$\lambda_i(t) = \frac{c_j'(t)}{v+c_j(t)}.$$

Integrating up the hazard rate gives

$$Q_i(t) = 1 - \frac{v+c_j(0)}{v+c_j(t)}[1-Q_i(0)].$$

If $c_j'(t)$ is higher while $c_j(t)$ is the same, then i quits at a faster rate, implying that it can build its reputation faster. However, if j keeps its derivative the same but increases its level (i.e., $c_j(t)$), then state i needs to quit at a slower rate, making i slower at building its reputation.

Suppose for now that $c_i(t) = tc_i + d_i$, where $c_i, d_i > 0$ for $i = 1, 2$. This simple example is sufficient to make some important comparative statics analyses. The term d_i could measure the sunk costs that are financially costly ex-ante (such as mobilizing troops), while c_i could measure freedom of the press (the idea being that with a freer press, conflict escalation translates more quickly into political pressure). Similar arguments in the proof of Lemma 1 implies that in equilibrium, state 1 will not quit after time

$$t_2^* = min\left\{\frac{w-d_2}{c_2}, \frac{(1-z)(v+d_1)}{zc_1}\right\}.$$

The first ratio is the time t beyond which state 2 prefers to attack (i.e., $c_2(t) = w$). The second term is the time t that state 2's reputation reaches 1 (i.e., $Q_2(t) = 1-z$). Likewise, state 2 will not quit after time

$$t_1^* = min\left\{\frac{w-d_1}{c_1}, \frac{(1-z)(v+d_2)}{zc_2}\right\}.$$

Hence, the horizon of the crisis bargaining game will be $t^* = \min\{t_1^*, t_2^*\}$, and state i will be advantaged if and only if $t_i^* < t_j^*$ (Corollary 1).

Suppose first that $c_i = c$ for $i = 1, 2$, but $d_1 > d_2$ so that the states' derivatives of audience costs are the same, but state 1 always has a higher level of audience costs. Since $\frac{w-d_1}{c} < \frac{w-d_2}{c}$ and $\frac{(1-z)(v+d_2)}{zc} < \frac{(1-z)(v+d_1)}{zc}$ hold, we have $t_1^* < t_2^*$. Thus, the following result immediately follows.

Corollary 2. *Suppose that $c_i(t) = tc + d_i$, where $c > 0$ and $d_1 > d_2 > 0$. State 1 is advantaged in the crisis bargaining game G regardless of the values of $v, w, z,$ or c.*

On the other hand, if we let $c_1 > c_2$ but $d_i = d > 0$ for $i = 1, 2$, then we achieve a result similar to Proposition 2. That is, state 1 is advantaged if and only if $z < \frac{v+d}{v+w}$. Hence, we can conclude that a higher level of audience costs is always advantageous, but a higher derivative of audience costs is not.

5. Concluding Remarks

The primary message of the model is that the audience costs mechanism is not always helping leaders to attain diplomatic success. Contrary to conventional wisdom, the ability of generating greater audience costs may be good or bad depending on (1) the benefit-cost ratio of the crisis (i.e., v/w); (2) initial probability of resolve, z; and (3) the states' sensitivities to audience costs (i.e., $c_i(t)$).

In Fearon's setup, the democratic state—the one that can generate higher audience costs—always has a reputational advantage because in equilibrium, the probability of facing a resolved opponent is always lower (higher) for the democratic (the autocratic) state. This reputational superiority of the democratic state compensates the drawbacks of having greater sensitivity to audience costs. The current model destroys the democratic state's reputational privilege by allowing a setup in which states' initial probability of resolve is independent of their costs of war or valuations for the prize.

In the current model, increasing a player's audience costs may increase or decrease his own payoff, depending on how the equilibrium horizon (i.e., the time beyond which no type quits) is determined. In particular, if the horizon is determined as in the model of Fearon (i.e., the horizon is the time at which audience costs are so high that all types prefer attacking over backing down), then increasing audience costs increases payoffs, for much the same reason as in Fearon (1994). If instead the horizon is determined as in the work of Abreu and Gul (i.e., the horizon is the time at which players become certain that the opponent is the resolved type), then increasing audience costs decreases payoffs, for much the same reason as increasing delay costs decreases payoffs in Abreu and Gul (2000).[14] Therefore, this paper shows that the Fearon regime obtains when the benefit-cost ratio of war is high or the initial probability of resolve is low, whereas the Abreu and Gul regime obtains when the benefit-cost ratio is low or when the initial probability of resolve is high.

Furthermore, analyses suggest that how a state generates its audience costs may play an important role in its diplomatic success. Generating higher levels of audience costs (i.e., higher $c_i(t)$) is always good for a state given that the state can keep $c_i'(t)$ low. However, generating higher levels of audience costs with high derivative may be bad—depending on the benefit-cost ratio of the crisis.

6. References

[1] Abreu, D., and F. Gul, (2000): "Bargaining and Reputation," *Econometrica*, 68, 85-117.

[2] Abreu, D., and R. Sethi, (2003): "Evolutionary Stability in a Reputational Model of Bargaining," *Games and Economic Behavior*, 44, 195-216.

[3] Eyerman, J., and R. A. Hart Jr., (1996): "An Empirical Test of the Audience Cost Proposition," *Journal of Conflict Resolution*, 40(4), 597-616.

[4] Fearon, D. J., (1994): "Domestic Political Audiences and the Escalation of International Disputes," *American Political Science Review*, 88, (September), 577-592.

[5] Gelpi, C. F., and M. Griesdorf, (2001): "Winners or Losers: Democracies in International Crises, 1918-1994," *American Political Science Review*, 95(3), 633-47.

[6] Guisinger, A., and A. Smith, (2002): "Honest Threats," *Journal of Conflict Resolution*, 46(2), 175-200.

[14] According to Abreu and Gul (2000), increasing my delay costs increases your equilibrium concession rate, and the condition that our reputations must hit 1 simultaneously implies that I must concede initially with higher probability.

[7] Harsanyi, J. C., (1967): "Games with Incomplete Information Played by 'Bayesian' Players, I-III," *Management Science*, 14, 159-182.

[8] Hendricks, K., A. Weiss, and R. Wilson (1988): "The War of Attrition in Continuous-Time with Complete Information," *International Economic Review*, 29, 663-680.

[9] Iida, K. (1993): "When and How Do Domestic Constraints Matte? Two-Level Games with Uncertainty," *The Journal of Conflict Resolution*, 37, 403-426.

[10] Kambe, S. (1999): "Bargaining with Imperfect Commitment," *Games and Economic Behavior*, 28, 217-237.

[11] Kreps, D. M., and R. Wilson (1982): "Reputation and Imperfect Information," *Journal of Economic Theory*, 27, 280-312.

[12] Leventoglu, B., and A. Tarar, (2005): "Prenegotiation Public Commitment in Domestic and International Bargaining," *American Political Science Review*, 99(3), 419-433.

[13] Milgrom, P., and J. Roberts (1982): "Predation, Reputation, and Entry Deterrence," *Journal of Economic Theory*, 27, 280-312.

[14] Myerson, R. (1991): *Game Theory: Analysis of Conflict.* Cambridge, MA: Harvard University Press.

[15] Partell, P. J., and G. Palmer, (1999), "Audience Costs and Interstate Crises: An Empirical Assessment of Fearons Model of Dispute Outcomes," *International Studies Quarterly*, 43(2), 389-405.

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

[17] Sartori, A. E., (2002): "The Might of the Pen: A Reputational Theory of Communication in International Disputes," *International Organization*, 56 (February), 121-149.

[18] Schelling, T. (1966): "Arms and Influence," Yale University Press.

[19] Schultz, K. A., (2001): "Looking for Audience Costs," *Journal of Conflict Resolution*, 45, 32-60.

[20] Schultz, K. A., (2012): "Why We Needed Audience Costs and What We Need Now", *Security Studies*, 21, 369-375.

[21] Slantchev, B., (2006): "Politicians, the Media, and Domestic Audience Costs," *International Studies Quarterly*, 50, 445-77.

[22] Snyder, J., and E. D. Borghard, (2011): "The Cost of Empty Threats: A Penny, Not a Pound," *American Political Science Review*, 105, 437-456.

[23] Smith, A., (1998): "International Crises and Domestic Politics," *American Political Science Review*, 92, 623-638.

[24] Tomz, M., (2007): "Domestic Audience Costs in International Relations: An Experimental Approach," *International Organization*, 61, 821-840.

[25] Trachtenberg, M., (2012): "Audience Costs: An Historical Analysis," *Security Studies*, 21, 3-42,