Is There a Trade-Off between Audience Costs and Diplomatic Success in Crisis Bargaining?

Selçuk Özyurt*
Sabancı University

Abstract

It is conventional wisdom in crisis bargaining literature that the ability of generating higher audience costs is an advantage. 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 resolution, and (3) states’ sensitivities to audience costs. In particular, if the value of the prize over the cost of attacking is low or the initial probability of resolution is high enough, then having greater sensitivity to audience costs undermines democratic states’ diplomatic success.

*Faculty of Arts and Social Sciences, Sabancı University, 34956, Istanbul, Turkey. (e-mail: ozyurt@sabanciuniv.edu)
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 on average 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.\(^1\) The audience costs may occur if the leader makes public threats or commitments but fails to carry through on them. It helps 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 (i.e, audience costs coefficient) and this is true regardless of the states’ resolution (cost of war), value of the prize, or the uncertainty about the type of opponent. Fearon explicitly states this point in his paper in several occasions and adds (page 585): “... it 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 is bad.\(^2\)

For this purpose, I use the crisis bargaining game of Fearon (1994) with no structural change, except the source of uncertainty. Fearon interprets the states’ costs of war as their level of resolve and thus models uncertainty about the states’ resolution by assuming that each state’s cost of war is a random variable to its opponent. However, I show that a different interpretation of “resolution” could lead to technically easier equilibrium, that upsets the dynamics of the Fearon’s model. Fearon’s interpretation of resolution implies that a state is resolved (i.e, not willing to back down) if and only if its war cost is sufficiently low. However, I suggest that resolution does not have to go hand in hand with cost of war. Whatever the underlying reason of the state’s resolution is, I suppose that a state is either resolute or not. Therefore, in my model, a state may be resolved even if it has a very high cost of war.

In what follows, a resolute leader is stubborn and committed not to back down from his demand(s). This does not necessarily imply that a resolute leader is irrational. Rational explanations for resoluteness and further discussions about the model are provided in the following section. Thus, in the spirit of Harsanyi (1967), I assume that each state has two types: “\textit{resolute}” and “\textit{flexible}” (not resolute). The flexible type has no commitment to his demand and is willing to back down and mimic the resolute type (i.e, bluff) if optimal. This interpretation of resolution is consistent with that of Schelling (1966): he suggests that states may be resolute or not and that they wish to acquire reputations for resolution in order to increase their credibility in international conflicts.

As the state’s type is not certain, there are two commitment devices that flexible players can utilize 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 opportunity of mimicking the resolute type and building false reputation on resolution. If the “\textit{benefit-cost ratio of the crisis}” is high—which is the case when the value of peaceful resolution is high, or the cost of military intervention is low—or the initial probability of resolution is low enough, then the former commitment device is more effective. In that case, the threat of costly war gives the advantage to the state that is more sensitive to audience costs (call it \textit{democratic state}). However, if the benefit-cost ratio of the crisis is sufficiently low or the initial probability of resolution is high enough, then it is too risky to use attacking

\(^2\)Schultz (2012) supports this argument with no formal treatment.
as a credible threat. Thus, the second commitment device—building false reputation on resolution by mimicking the resolute type—becomes more effective. In this case, the second state that is less sensitive to audience costs (call it autocratic state) has the advantage.

Section 2 explains the details of the continuous-time crisis bargaining game. Section 3 provides the main results of this paper. In the last part of Section 3, 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 sensitivity to audience costs. In particular, I show in Section 4 that a higher level of audience costs is always good, but a higher sensitivity to audience costs is not. Finally, I conclude in Section 5.

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 expected benefit net of losses due to attacking. 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. In case of simultaneous concessions, the prize is divided equally.\(^3\) The values of all these parameters are common knowledge.

\(^3\) 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 \( \frac{1}{2} - c_i(t) \). Finally, if states escalate the conflict indefinitely, each state gets a payoff that is strictly less than its war payoff. These particular assumptions are not crucial because in equilibrium, simultaneous concessions or attacks or escalations with infinite horizon occur with probability 0.
Finally, I assume that there is some uncertainty about states’ resolution. In the spirit of Harsanyi (1967), each state has two types: resolute or flexible. The prior probability that state $i$ is resolute is denoted by $z_i \in (0, 1)$. Each state knows its own type but does not know the opponent’s true type. Call this crisis game, where all parameters are common knowledge, $G$.

The flexible types pick a strategy, given their beliefs, to maximize their expected payoff in the crisis game. On the other hand, a resolute type never backs down. I also assume that the resolute types understand the equilibrium and attack immediately once they are convinced that their opponent will never quit. Fearon’s resolute types, in fact, satisfy these two assumptions with one negligible deviation. With these two assumptions, I simply suppose that the resolute types prefer a peaceful resolution to war, but backing down is not an option for them. A resolute state would be loving war (i.e., $-w_i \geq 0$), having very high audience costs, or thinking of not weakening its reputation for future negotiations. According to Fearon (1994), a state is resolute if and only if it’s cost of war is sufficiently low. However, in my model, whatever the underlying reason of the state’s resolution is, a state is either resolute or not (i.e., flexible). Therefore, in contrast to Fearon (1994), I simply allow the possibility that a state with a high cost of war would be resolute.

On the other hand, a leader who is engaged in a dispute with another state may be resolute 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. Resolute 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). 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) supports 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

---

4In particular, Fearon’s resolute types understand the equilibrium and attack after they are convinced that their opponent will never quit. The term immediate in my assumption does not change the flexible players’ equilibrium behaviour. It simply, and only, ensures that the game will finish when players stop quitting. Thus, eliminating the term immediate in my assumption would not change the results in the following sections.

5These “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. Resolute types that I define in this paper are closer to the obstinate types of Atakan and Ekmecki (2013) and of Özyurt (2013). In both of these papers, the obstinate types understand equilibrium and quit negotiation once they are convinced that their opponents are also obstinate.
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).

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 resolution, 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.

**Strategies**

Strategies of resolute players are simple: never back down and attack when convinced that the opponent will never quit. On the other hand, flexible types have the opportunity of credibly bluffing, manipulating the adversaries’ belief, and thus building false 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 $\sigma = (Q, A)$ of state $i$ is defined by a collection of cumulative distributions. $Q_i(t) : \mathbb{R} \cup \{\infty\} \rightarrow [0, 1]$ represents the probability that state $i$ quits by time $t$ (inclusive), and $A_i(t) : \mathbb{R} \cup \{\infty\} \rightarrow [0, 1]$ denotes the probability that state $i$ attacks by time $t$ (inclusive). Note that the distribution functions $Q_i(t)$ and $A_i(t)$ are state $j$’s belief about
state $i$’s play. It follows that $Q_i(t) + A_i(t) \leq 1$ for all $t \geq 0$ and that
\[
\lim_{t \to \infty} Q_i(t) \leq 1 - z_i.
\]

**Remark 1.** Note that $Q_i(t)$ gives the probability $i$ will quit prior to $t$. The probability that $i$ will quit prior to $t$ given that he is flexible is higher, which is equal to $Q_i(t)/(1 - z)$.

Given $Q_j$ and $A_j$, flexible state $i$’s expected payoff of quitting at time $t$ is
\[
U_i(t, Q_j, A_j) := v_i Q_j(t) - w_i A_j(t) + [1 - Q_j(t) - A_j(t)] [-c_i(t)]
\]
\[+ \left( \frac{v_i}{2} - tc_i \right) [Q_j(t) - Q_j(t^-)] \quad (1)
\]
with $Q_j(t^-) = \lim_{y \to t} Q_j(y)$.

### 3. Main Results

In this section, I analyze two special cases. These particular cases both convey the flavor of the analysis and are furthermore 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 for now that both states are known to be flexible (i.e., $z = 0$). In this case, higher sensitivity to audience costs is always an advantage for diplomatic success. In any subgame perfect equilibrium, state $i$ does not back down beyond time $t_i$, where $i$’s audience costs is 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 $t_1 < t_2$. In equilibrium, state 1 anticipates that if he waits until time $t_1$, state 2 would certainly back down prior to $t_1$ (i.e., there is no point for state 2 to escalate the crisis beyond $t_1$ or to attack prior to $t_1$). Thus, state 1 never quits. Finally, in equilibrium, 2 anticipates that delaying the concession has no benefit and so quits at time 0. Hence, the conflict is resolved before it escalates, and payoffs of states 1 and 2 are $v$ and 0, respectively.

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 its reputation.
on its resolution by bluffing and escalating the crisis. The (sequential) equilibrium of
the crisis game is unique. A notable implication of the equilibrium is that sensitivity to
audience costs is not always an advantage for diplomatic success.

In a sequential equilibrium, 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,
estimation 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.

We describe state \( i \)'s behavior in the crisis game by a probability distribution over
quitting times, \( Q_i(t) = Pr(i \text{ will quit prior to } t) \), where 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. That is, \( \lambda_i(t) = \frac{dQ_i(t)/dt}{1-Q_i(t)} \). We look for an equilibrium where flexible player \( j \)
mixes between quitting and escalating. Therefore, flexible \( j \) is indifferent between quitting
at time \( t \) and waiting for an infinitesimal period \( \Delta \) 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 \( \Delta \).

Solving this equation for \( \lambda_i(t) \) and taking its limit as \( \Delta \) approaches zero yields

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

Integrating up the hazard rate gives

\[
Q_i(t) = 1 - \frac{va_i}{v + tc_j},
\]

where \( a_i = 1 - Q_i(0) \).

**Proposition 1.** A sequential equilibrium of the crisis bargaining game \( G \) is characterized
by the following conditions: For \( i = 1, 2 \),

1. \( Q_i(t) = 1 - \frac{va_i}{v + tc_j} \) for all \( t \leq t^* \),

2. \( a_i \in [0, 1] \) and \( [1 - a_1][1 - a_2] = 0 \),

\( ^6 \) 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 \( t + \Delta \).
3. $t^*$ solves $Q_2(t^*) = 1 - z$ and $Q_1(t^*) \leq 1 - z$, and

4. $A_i(t) = 0$ for all $t < t^*$ and $A_i(t) = 1 - Q_i(t^*)$ for all $t \geq t^*$.

Proof. (Sketch) The proof proceeds by observing that any sequential equilibrium pair $(Q_1, A_1), (Q_2, A_2)$ must have the following properties.

(i) Flexible player $i$ will not hesitate to quit or attack at time $t$ once he knows his opponent is resolute or $t \geq w/c_j$. Likewise, a resolute player will attack once he knows his opponent is resolute. Thus, there is some $t^* \leq w/c_1 < w/c_2$ such that

$$Q_i(t) < 1 - z$$

for all $t < t^*$ and

$$Q_i(t^*) = 1 - z \text{ for } i = 1, 2 \text{ if } t^* < w/c_1$$

$$Q_1(t^*) \leq 1 - z \text{ and } Q_2(t^*) = 1 - z \text{ if } t^* = w/c_1$$

(ii) If $Q_i$ jumps at $t$, then $Q_j$ is constant at $t$. The reason is that $j$ would always want to wait $dt$ amount of time in order to enjoy the discrete chance of $i$ quitting.

(iii) If $Q_j$ 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 $t'' < t^*$ 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.

From (i) – (iv) it follows that $Q_1$ and $Q_2$ will be continuous and strictly increasing on $[0, t^*]$. But if both are quitting with positive probability, then they must be quitting at hazard rates $\lambda_1, \lambda_2$ as defined above, so

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

as defined above. By (ii) both of $Q_1(0)$ and $Q_2(0)$ cannot be positive, implying that $\min[1-a_1, 1-a_2] = 0$. By (i) we know that $t^*$ satisfies $Q_2(t^*) = 1 - z$ and $Q_1(t^*) \leq 1 - z$. Finally, the fact that $Q_i(t) < 1 - z$ for all $t < t^* \leq w/c_1$, states will not attack before time $t^*$.

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 instantaneous quitting 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, if \( Q_1(0) = Q_2(0) = 0 \), then with the condition that the crisis game G reaches time \( t > 0 \), the probability that state 2 is resolute will be greater than the probability that state 1 is resolute. The last observation implies that under certain circumstances, state 2 can build its reputation much faster, and thus, it is easier for state 2 to convince the adversary (about its resolution) if it bluffs.

The following two lemmas solve the equilibrium horizon of the crisis game \( t^* \) as a function of the primitives and find the equilibrium strategies \( Q_1(t) \) and \( Q_2(t) \).

**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 \( i, j \in \{1, 2\} \) and \( j \neq i \).

**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 zero with a positive probability (i.e. \( Q_j(0) = 0 \)), state \( i \) will be convinced that \( j \) is resolute by time \( \tau_j \), solving \( Q_j(\tau_j) = 1 - \frac{w}{v+\tau_j c_i} = 1 - z \), 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 resolute no later than \( \tau_j \). Hence, state \( i \) never backs down after time \( t_j^* := \min \left\{ \frac{w}{c_j}, \frac{v(1-z)}{zc_j} \right\} \). Similar arguments hold for state \( j \), i.e. 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. The reason for this is simple. If \( t^*_1 = \frac{w}{c_1} \), then the optimal strategy for flexible 2 is to quit immediately to escape from a possible war (note that \( c_2 t^*_1 < w \) since \( t^*_1 < t^*_2 \)). However, if \( t^*_1 = \frac{v(1-z)}{zc_2} \), then 2 will be convinced that 1 is resolute, and so 1 will never quit. Once again, flexible 2’s optimal strategy is to quit immediately at time \( t^*_1 \) and avoid war. As a result, the minimum of \( t^*_1 \) and \( t^*_2 \) determines the end of the escalation. Hence, in equilibrium, given that flexible players randomize the timing of quitting, escalation continues until time \( \min\{t^*_1, t^*_2\} = t^* \) with some positive probability and stops at this time with certainty. \[\square\]
Lemma 2. If $t^*_i > t^*_j$, where $i, j \in \{1, 2\}$ and $i \neq j$, then sequential equilibrium strategies are uniquely determined as follows:

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

where

$$a_i = \begin{cases} z + \frac{c_j z (1 - z)}{v}, & \text{if } \frac{w}{c_j} > \frac{v(1 - z)}{zc_i} \\ z + \frac{cw}{v}, & \text{otherwise.} \end{cases}$$

Proof. Since $t^*_i > t^*_j$, we have $t^* = t^*_j$. Therefore, $\frac{v(1 - z)}{zc_i} = \tau_i > t^*$ if $Q_i(0) = 0$. Since the game ends before time $\tau_i$ 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 $Q_j(0) = 0$.

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

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 $t^*$, flexible states are indifferent between quitting at time 0 and waiting for some time $t < t^*$ and then quitting at this time. Moreover, according to Proposition 1, $A_i(t) = 0$ for all $t < t^*$ and $i = 1, 2$. Therefore, flexible state $i$’s equilibrium payoff in the crisis bargaining game is

$$U_i = vQ_j(0) + [-c_i(0)](1 - Q_j(0)) = vQ_j(0),$$

which is given by Equation (1). The second condition of Proposition 1 implies that in equilibrium, $Q_1(0)Q_2(0) = 0$ must hold. Thus, given the parameter values, only 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 $U_j = vQ_j(0) > 0$. 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 $t^*_i > t^*_j$. 

11
**Proposition 2.** State 1—the one which is more sensitive to audience costs—is advantaged in the crisis bargaining game G if and only if the benefit-cost ratio of the crisis is large enough or the initial probability of resolution is low enough (i.e., $z < \frac{v}{w + c_1}$).

**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 $\tau_1 = \frac{v(1-z)}{zc_1} > \tau_2 = \frac{v(1-z)}{zc_2}$. 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.

**Remark 2.** Note that State 2—the one which is less sensitive to audience costs—is advantaged in the crisis bargaining game G whenever the benefit-cost ratio of the crisis is small or the initial probability of resolution is high enough (i.e., $z > \frac{v}{w + c_1}$).

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, at any time $t < t^*$, it is more likely that the advantaged state will get the prize. 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 $z < \frac{v}{w + c_1}$, 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 $z > \frac{v}{w + c_1}$, 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 state 1’s sensitivity to audience costs 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 $Q_2(0) = 1 - a_2$, which is equal to $1 - z - \frac{w}{v}$ 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

\[ q_A = \max \left\{ 1 - \frac{w_c_2}{(1 - z)(w_c_1 + w_c_2)}, 0 \right\} \]

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 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.

A final observation in this section is consistent with the idea of democratic peace. Although the idea of democratic peace has circulated since the time of Immanuel Kant, it was not scientifically evaluated until the 1960s. It relies on one of the most thoroughly tested observations in international politics, that democracies (for some appropriate definition of democracy) rarely, or even never, go to war with one another. As it is easy to verify, \( q_A \) (weakly) decreases as \( c_1 \) and \( c_2 \) take values closer to each other. That is, war is less likely if the states’ sensitivities to audience costs are closer to one another.

**Comparison with Fearon’s Model**

In this paper, I start with an assumption that a player is either a commitment/resolved type (i.e., not willing to quit) or a flexible type (i.e., willing to quit). The commitment type would be loving war, valuing the prize highly, having a very high audience cost coefficient, or thinking of not weakening its reputation by backing down. Whatever the reason is, a state is either a resolute type or a flexible type. Fearon’s interpretation of resolution implies that only the types with low cost of war could be the commitment type. However, my interpretation of resolution allows the idea that even a state with a very low cost of war could be the flexible type and a state with a very high cost of war could be the commitment type. This fundamental difference in the interpretation of resolution, though it may seem as if it is a nuance in technical assumptions, is the main source of the different results of these two papers.

The crisis bargaining game \( G \) in the current paper is identical to Fearon’s original game except for the source of uncertainty. Fearon (1994) assumes that each state is

\[ \text{For further discussion, see, for example, Lipson (2003), Schultz (1999), and Russett (1995).} \]
uncertain about the opponent’s cost of war but knows its own. Therefore, states’ beliefs are represented by a distribution function $F$ over the set of types $[w, 0]$, where $w < 0$. Fearon shows, in Lemma 2, that in equilibrium with horizon $t^*$, the set of types is divided into three classes.

(i) All the types $w_i \in [w, 0]$ satisfying $w_i > -c_i(t^*)$ escalate the conflict up to time $t^*$ and then attack. These types never quit.

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

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

Therefore, in equilibrium, audience costs separate the players (types) according to their willingness to fight. In equilibrium with horizon $t^* > 0$, the probability that state 1 follows a strategy involving attack is equal to $1 - F(-c_1(t^*))$ (i.e., the probability that $w_i \geq -c_i(t^*)$). Likewise, the probability that state 2 follows a strategy involving attack is equal to $1 - F(-c_2(t^*))$. 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, state 1 attacks at time $t^*$ with a greater probability than state 2. In equilibrium, state 2 can anticipate this and quits before $t^*$ with a greater probability. This makes state 1 advantaged regardless of the primitives.

To be able to draw analogies between Fearon’s model and the current one, I suppose for now that the cost of war of the states (in my model) is some $w < 0$. In equilibrium with horizon $t^* > 0$, if the cost of war is high (i.e., $w < 0$ is low) enough, then we will have $w < -c_1(t^*) < -c_2(t^*)$. In fact, we will have $t^* = v(1 - z)/zc_1$ when $i$’s audience costs satisfy $c_i(t) = c_it$. In this case, flexible types of both states will certainly quit before time $t^*$. Therefore, the probability that state 1 (or 2) follows a strategy involving attack is equal to $z$. Note that $z$ is the probability that a state is resolute, and thus, it is the lower boundary for the probability of a state attacking in equilibrium. Since both states’ probabilities of attacking at time $t^*$ will be the same, state 1 will lose its favorable position. Because of this, the contest between the states will turn into a “Which state will quit first?” contest. Or equivalently, the state that persuades the other state about its resolution first will win the competition. Since state 1 is more sensitive to audience

---

8Since both states’ probabilities of attacking at time $t^*$ will be the same, the contest becomes a “Which state will persuade the other state about its resolution first?” contest. State 2 can build its reputation faster given that state 1 does not quit at time 0 with positive probability (i.e., $Q_1(0) = 0$). Therefore, in equilibrium, state 1 will play a strategy $Q_1(t)$ such that $Q_1(0) > 0$, making 2 advantaged.
costs, quitting after \( t = 0 \) is more costly for it. Therefore, state 1 quits with a smaller rate beyond \( t = 0 \) (recall that the hazard rates satisfy \( \lambda_1(t) < \lambda_2(t) \) for all \( t \)). In equilibrium, state 1 anticipates this situation and thus quits at time \( t = 0 \) with a positive probability, making state 2 advantaged. This case never occurs in Fearon’s model. This is because in the model of Fearon (1994), there is no lower boundary on the probabilities that states would attack, and thus, the probability that state 1 will attack at time \( t^* \) will always be greater than the probability that state 2 will attack at time \( t^* \).

When the cost of attacking is small (i.e., \( w < 0 \) is high) enough, then my model will generate dynamics that are similar to Fearon’s original model. In equilibrium with escalation \( t^* > 0 \), if the cost of war is low enough, then we will have \( w = -c_1(t^*) < -c_2(t^*) \). In fact, we have \( t^* = -w/c_1 \) since \( i \)’s audience costs is \( c_i(t) = c_i t \). In this case, flexible state 1 will quit before \( t^* \) and attack at time \( t^* \) with positive probabilities. Therefore, the probability that state 1 attacks at time \( t^* \) is equal to \( z + [1 - Q_1(t^*)/(1-z)] \), where \( 1 - Q_1(t^*)/(1-z) > 0 \) denotes the probability that flexible state 1 attacks at time \( t^* \). However, the probability that state 2 follows a strategy involving attack is equal to \( z \). Since, state 1 is more likely to attack at time \( t^* \), state 2 can anticipate this in equilibrium and quits at the early stages of the escalation with a greater probability. This makes state 1 advantaged.

4. Some Extensions

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 level of audience costs (i.e., \( c_i(t) \)) and sensitivity to audience costs (i.e., \( c_i'(t) \)). \(^9\) 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 sensitivity low, then this must be good. However, generating audience costs simply by producing greater sensitivity would be bad. In this section, I will slightly extend the model to clarify the distinct impacts of the level of audience costs and the sensitivity to 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.

\(^9\) I am grateful to the referees whose comments shaped the discussion in this section greatly.
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 $\Delta$ 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 $j$ is more sensitive to audience costs (i.e., $c_j'(t)$ 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 sensitivity 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 $c_i(t) = c_i t + d_i$, where $c_i, d_i > 0$ for $i = 1, 2$. This simple example is sufficient to make some important comparative statics analyses. 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 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’ sensitivities to 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 $c_i(t) = ct + d_i$, where $c > 0$ and $d_1 > d_2 > 0$. State 1 is advantaged in the crisis bargaining game $G$ regardless of the value of $v, w, z$, or $c$. 

16
On the other hand, if we let $c_1 > c_2$ but $d_1 = d > 0$ 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 higher sensitivity to audience costs is not.

5. Concluding Remarks

The primary message of the model is that the audience cost mechanism is not always helping leaders 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 resolution, $z$, and (3) states’ sensitivities to audience costs (i.e., $c_i(t)$).

Furthermore, analyses suggest that the consequences of higher audience costs are not delivered in a single packet. Generating higher levels of audience costs (i.e., higher $c_i(t)$) is always good for a state given that the state can keep its sensitivity to audience costs low. However, generating higher levels of audience costs with high sensitivity is bad if the benefit-cost ratio of the crisis is low or if the state’s initial probability of resolution is high. A state may increase its sensitivity to audience costs by, for example, making hostile and aggressive threats/statements, possibly tying the conflict with some ethnic or religious disagreements. However, a state may generate greater levels of audience costs with lower sensitivity by taking financially costly mobilization or arming measures.

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 that the democratic state attacks is always higher. This reputational superiority compensates any drawback 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 resolution is independent of their costs of war. In that case, when states’ costs of war are high enough, the negative consequences of audience costs (sensitivity) will be eminent because states will accumulate audience costs faster, making escalation and a long equilibrium horizon riskier.

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 a resolute type), then increasing audience costs decreases payoffs, for much the same reason as increasing delay costs decreases payoffs in Abreu and Gul (2000). Therefore, this paper shows that the Fearon regime obtains when the benefit-cost ratio of war is high or the initial probability of resolution is low, while the Abreu and Gul regime obtains when the benefit-cost ratio is low or when the initial commitment probabilities are high.

The main message of the current paper promotes various policy implications. If a democratic state is facing an autocratic state that values the prize highly and the cost of war is very high for all sides, then the democratic state can or should decrease its cost of military intervention by eliminating some of the risks involved in such interventions. Another strategy that is likely helping democratic states in a conflict with a low benefit-cost ratio is to increase their value of the prize by convincing their domestic audiences that the value of the good in dispute is high.

Although the current model takes the values for the prize $v$, cost of war $w$, and sensitivity to audience costs $c'_i(t)$ as given, in reality, states may have the power to change these parameters. An important conjecture directly implied by the equilibrium analysis is that increasing its sensitivity for audience costs by sending/making strong threats or commitments may be a bad policy if the benefit-cost ratio of the crisis is not sufficiently high. On the other hand, a lengthier escalation could facilitate diplomatic success for democratic states because escalation may help them to buy some time to change the benefit-cost ratio of the crisis. Since democratic states can increase the likelihood of a lengthier escalation by reducing their sensitivity to audience costs, following a passive or reserved management during a crisis where the benefit-cost ratio is low may be optimal.

---

10 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.
Appendix

Proof of Proposition 1. For \( i = 1, 2 \), let \( t_i = \inf \{ t \geq 0 | Q_i(t) = \lim_{k \to \infty} Q_i(k) \} \) where \( \inf \emptyset := \infty \). That is, \( t_i \) denotes the time that state \( i \)'s reputation reaches one. Also, let \( \hat{t}_i = \frac{w}{c_i} \) denote the time that state \( i \) is indifferent between backing down and attacking. Thus, in equilibrium, flexible states finish the game prior to time \( t^* = \min \{ t_1, t_2, \hat{t}_1, \hat{t}_2 \} \). Since backing down is more beneficial than attacking for both states until time \( t^* \), we must have \( A_i(t) = 0 \) for all \( t < t^* \). Proofs of the following arguments directly follow from Hendricks, Weiss and Wilson (1988) and are analogous to the arguments in the proof of Proposition 1 in Abreu and Gul (2000). Therefore, I skip the details.

Lemma A.1. If state \( i \)'s strategy \( Q_i \) is constant on some interval \( [t_1, t_2] \subseteq [0, t^*) \), then state \( j \)'s strategy \( Q_j \) (where \( j \neq i \)) is constant over the interval \( [t_1, t_2 + \eta] \) for some \( \eta > 0 \).

Lemma A.2. For any state \( i \), \( Q_i \) does not have a mass point over \( (0, t^*) \).

Lemma A.3. \( Q_1(0)Q_2(0) = 0 \).

Therefore, according to Lemma A.1 and A.2, \( Q_i \) is strictly increasing and continuous over \([0, t^*)\). Therefore, the utility function of state \( i \) given in Equation (1) is also continuous on \([0, t^*)\). Then, it follows that \( D^i := \{ t | U_i(t, Q_j, A_j) = \max_{s \in [0, t^*]} U_i(s, Q_j, A_j) \} \) is dense in \([0, t^*)\). Hence, \( U_i(t, Q_j, A_j) \) is constant for all \( t \in [0, t^*) \). Consequently, \( D^i = [0, t^*) \). Therefore, \( U_i(t, Q_j, A_j) \) is differentiable as a function of \( t \). Similarly, \( Q_i \) is differentiable because the utility function is differentiable on \([0, t^*)\). Since \( j \) is indifferent between waiting and quitting at all time \( t \leq t^* \) we have \( U_j(t, Q_i, A_i) = k \) for some \( k \in \mathbb{R} \). Differentiating the both sides and rearranging yields \( \frac{dQ_i(t)}{dt} = \frac{c_j}{1-Q_i(t)} \). Integrating up the hazard rate gives \( Q_i(t) = 1 - \frac{e^{-ct_j}}{a_k} \) where \( a_k = 1 - Q_1(0) \). By Lemma A.3, we know that \( Q_1(0)Q_2(0) = 0 \) implying condition (ii). Optimality implies that \( Q_2(t^*) = 1 - z \) whereas \( Q_1(t^*) \leq 1 - z \) because it might be the case that \( c_1t^* = w_1 \). Therefore, \( A_2(t) = 1 - z \) and \( A_1(t) = 1 - Q_1(t^*) \) for all \( t \geq t^* \).

References


