

# Media Framing and Audience Costs

Peter D. Carey II

University of California, Merced

Melissa N. Baker

University of California, Merced

## **Abstract**

Work in international relations often assumes that a domestic audience is able to constrain and punish the actions of their leader. This work assumes that the audience perceives and understands the actions that their leaders take, and are able to hold them accountable. Implicitly, these works also assume that the audience receives an unbiased assessment of the leader's actions. Using an experiment, we show that the framing of information about a leader's actions has an impact on an individual's ability to punish their leader. We also use simulations to aggregate these attitudes and speak to attitudes of the mass public. When news of a leader's broken promise is framed negatively, the logic of audience costs holds and citizens punish the leader. When a broken promise is framed positively, however, citizens are less able to constrain their leader, resulting in a breakdown in the logic of audience costs.

# 1 Introduction

During the 1950s and 60s, the United States and the Soviet Union jockeyed for control over Berlin. Firmly situated in the Communist East, the Soviet Union repeatedly made overtures towards dissolving the United States' rights to the Western portion of Berlin. The response from the United States was unequivocal: in a televised speech to the nation on the 25th of July, 1961, President Kennedy told the nation that the US would not be moved by Soviet threats. President Kennedy clearly signaled to Premier Khrushchev that America could not and would not “negotiate with those who say ‘What’s mine is mine and what’s yours is negotiable’” (Trachtenberg 2012, p. 28).

This speech leads us to ask two questions. First, why would Kennedy signal his resolve about the Berlin situation using a televised speech? There were certainly more direct lines to the Kremlin, and he would not have needed to go through the trouble of making his threats public. Second, why would Khrushchev believe anything from Kennedy’s speech? It is entirely plausible that Khrushchev could see Kennedy’s threats as ‘cheap talk’ and expect the US to back down in the face of continued threats. A popular answer comes from the crisis bargaining literature, which suggests that Kennedy “went public” to generate audience costs (Fearon 1994a). These costs are the price a leader pays from making a promise to act and then backing down: the leader makes a promise, reneges, and then faces costs from their domestic audience, often in the form of being voted out of office. According to the theory, publicly announcing his intentions allowed Kennedy to send a believable signal to Khrushchev: I will not negotiate, because I have made it too politically costly to negotiate.

The theory expects that, because Kennedy is an elected leader who is sensitive to his domestic audience, Khrushchev should see Kennedy’s threat as believable and back down. Instead, what we observe is the construction of the Berlin Wall, and the US opening up negotiations in an effort to de-escalate the conflict. Meanwhile, at home, the American public

seemed relatively ambivalent to the whole adventure (Trachtenberg 2012). What happened? Why didn't audience costs force Kennedy to escalate in response to Khrushchev?

The ability of audience costs to constrain leaders is primarily dependent on the public's ability to learn about and understand the nature of the broken promise. In the ideal theoretical scenario, the public knows about the initial promise, receives a clear signal that it has been broken, and punishes their leader for renegeing. In practice, however, this information transmission is fuzzy, and the information itself is subject to distortion in a number of ways. The constraining effect of audience costs is dependent on how the domestic audience receives the information about the broken promise, which is often transmitted by the media. If the media portrays the broken promise in a negative light, the logic of audience costs holds. If, however, the media frames the broken promise as a good decision on the part of the leader, we expect the public to be less willing to constrain the leader. Trachtenberg (2012) echoes this sentiment for the Kennedy-Khrushchev example: "the administration was certainly not raked over the coals in the press for having moderated its position" (Trachtenberg 2012, p. 28). The President was able to back down precisely because the media was not raking him over the coals; rather, it framed the negotiations as something that "must go on, no matter how difficult and discouraging" (ibid.). In this case, we observe the media siding with, or at least refusing to condemn, the President's decision to renege on the deal, resulting in the public being ambivalent towards the broken promise.

In this paper, we experimentally demonstrate that framing does influence an individual's willingness to punish a leader for renegeing on promises. A negative frame will push respondent to punish the leader more harshly, while a positive frame reduces the likelihood of the leader being punished. These influences are important, because they form the basis for audience cost theory. If audiences are malleable, then there are conditions where a leader's word might not be as credible as audience cost theory predicts.

The paper proceeds as follows. In the next section, we outline the prominent literature

on audience cost theory and highlight why audience costs are important at the international level and what the literature says about the microfoundations of these costs. We then discuss the literature on framing effects, particularly on how framing can shape subsequent action. The theory then brings these two strands of literature together, demonstrating how framing effects shape the usefulness of audience costs as a signaling mechanism. The media can frame the leader's decision to break the promise in a negative light, making the public more likely to punish the leader and keep the logic of audience costs intact. However, the media can also frame the decision in a positive manner, which should make the public *less* likely to punish the leader. This means that leaders engaged in crisis bargaining must be aware of the media situation in the opposing country when observing threats from the leader: is the media friendly to the leader or hostile? If the media is friendly, then the leader's threats should not be treated as credible: the media will mitigate any potential audience costs the leader would have faced.

After outlining the theory, we describe the experimental design and results. We gave our subjects a vignette outlining a scenario in which the President made and then subsequently broke a promise. Depending on what condition subjects were in, they were given the straight 'facts of the case' with no attempt at a frame, a positive frame that mentioned how new information had come to light making it prudent to renege, or a negative frame that accused the leader of flip-flopping. As the audience cost theory suggests, a negative frame results in an increased willingness to constrain the leader for renegeing on their promise. A positive frame, on the other hand, is associated with a significant decrease in an individual's willingness to constrain the leader. This finding suggests that researchers must be wary of where and when they apply the logic of audience costs, as they are context-sensitive.

## 2 Audience Cost Literature

The idea of an ‘audience cost’ was first proposed by Fearon (1994a). Fearon wanted to know what some signals sent by leaders needed to be taken seriously while others could be safely ignored. This is particularly important in the context of international conflict, where ignoring a credible signal can result in a costly war, and respecting ‘cheap talk’ can leave a player suckered. Fearon focused on the nature of the statement: it gains credibility if made publicly. A leader runs the risk of being punished for backing down if they make a statement publicly and then change their minds, and opposing leaders know this, giving the threat credibility. This has proven to be an influential theory in the international relations literature, one with much macro-level empirical support<sup>1</sup>. This literature suggests that audience costs exists for all types of countries and regimes, but democracies are likely to be particularly sensitive to them because they are popularly elected. Thus, the original consensus was that audience costs only existed in democracies; this claim was later contested. Weeks (2008) suggests that audience costs can exist in autocracies, too, where members of the leader’s selectorate serve as the audience. Autocracies can thus also use public statements as a signal of resolve.

While not as extensively studied, there is also considerable interest in whether there is

---

<sup>1</sup>See, for example, C. R. Conrad, J. Conrad, and Young (2014), Davies and Johns (2013), Eyerman and Hart (1996), Fearon (1994b), Gibler and Hutchison (2013), Guisinger and Smith (2002), Kinne and Marinov (2012), Kurizaki and Whang (2015), McManus (2014), Partell and Palmer (1999), Potter and Baum (2010), Potter and Baum (2014), Prins (2003), Ramsay (2004), Schultz (2001), Smith (1998), Tarar and Leventoğlu (2009), Tarar and Leventoğlu (2012), Uzonyi, Souva, and Golder (2012), and Weeks (2008). Much of this work has been at the observational level: observing when a leader makes a public statement and noting the outcome. There have also been some studies looking at the microfoundations of audience costs. See, for example, Tomz (2007), which finds that respondents are sensitive to the act of breaking a promise. There is a nascent literature, which we add to, that suggests that there are limitations to audience costs. Kertzer and Brutger (2016), for example, find that audiences are less sensitive to the inconsistency of a leader breaking their promise and more sensitive to the belligerence of using force in the first place. Downes and Sechser (2012) suggest that there are flaws in the data that has been used up until this point, and their own observational data states that threats from democracies, which audience costs theory predicts should be more credible, are no more effective than threats from autocracies. These works, and ones like, point to major issues in both the microfoundations and macro-implications of audience cost theory.

micro-level support for audience costs. These authors examine the base assumptions that audience cost theory makes; namely, that the domestic audience is aware enough to impose costs on the leader. See Hoffman et al. (2009), Levy et al. (2015), Tingley and Walter (2011), Tomz (2007), and Trager and Vavreck (2011). This literature tends to look at how the audience thinks and how it responds to leaders reneging on promises. The general finding is that the audience values consistency in their leaders, and they punish for the inconsistency that backing down brings. They also value the national honor, and take steps to punish the leader for impugning on it.

It should be noted, however, that the support for audience costs is mixed, and it is attacked from all sides. One common critique is that “audience costs” only occur rarely, and thus they are not useful as concepts (Baum 2004; Snyder and Borghard 2011). Others argue that leaders are able to talk their way out of the promises made, allowing them to back down from a challenge without suffering the expected backlash (Levendusky and Horowitz 2012). Another area of concern lies with the audience itself. It could be that the audience does not act the way we expect it to; this is especially true if they have their own policy preferences (Chaudoin 2014). It could be that the audience does not receive the information like the theory suggests: neutrally, and without bias. The audience might also internalize the information in a different way, utilizing motivated reasoning to force the new information to square with their preconceived notions instead of rationally updating.

We focus on the audience as well, particularly in how they receive the information. We argue that, under certain situations, audience costs can exist, but we need to be cognizant of the media environment in the country that tries to employ public threats. Specifically, we need to understand how the domestic audience gets its information about the leader’s promises. If the transfer of information is biased, there are serious implications for whether audience cost theory holds.

## 2.1 The Logic of Audience Costs

Slantchev (2012) describes five key factors that have to be true in order for bluffing to be informative. First, backing down from the statement has to cause any costs previously incurred to increase. If this were not the case, any escalation would be “cheap talk.” Second, the costs need to increase as the crisis escalates; as leaders move up the chain of escalation, the costs associated with action need to be greater. Third, it needs to be possible that the costs can become large enough that war (or further escalation of the conflict) is preferable to concession. If not, any threat the actor makes could not be seen as credible: it would always be rational to back down in the face of war. Fourth, there must not be any other mechanism of coercion available to the escalating actor. Finally, attempting to coerce cannot increase the opponent’s cost of conceding. If coercion did influence the opponent’s cost of conceding, then we would expect the opponent to be less willing to back down following a threat.

If all five of these conditions are satisfied, two things are true. First, escalation can commit the escalating actor to fighting. This is because the cost of backing down becomes too great. Second, the risk of war discourages bluffing on the part of the escalating actor. These two factors allow the act of escalating a conflict to be informative to the opponent.

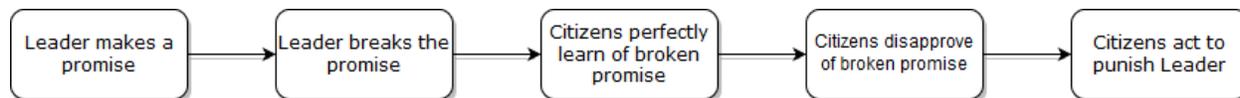
In less formal terms, the audience cost theory states that leaders can send informative signals about their level of resolve by making public threats. If the audience that observes the threat is willing to punish the leader for backing down, the threat is credible. This is because there is the legitimate possibility that the leader will face negative consequences for renegeing on their promise.

Weeks (2008) points out that there are three important factors that are required for leaders to generate audience costs. First, the audience has to have the “means and incentives” (Weeks 2008, p. 37) to punish the leader. These domestic actors must also see backing down after the threat as worse than conceding without the threat in the first place; this is similar to Slantchev’s second and third points. Finally, the possibility of the domestic costs need to

be observed by the actor that is receiving the threat.

The following diagram represents the main logic of audience costs. First, the leader makes some sort of promise to act under certain conditions. Referring to the Berlin example above, Kennedy promised that he would not negotiate with the USSR unless they took a less hardline position on America’s rights in Germany. The next step sees the leader breaking their promise. Here, we observe the Russians building the Berlin Wall and the US Administration agreeing to negotiations. The next step is that the citizens perfectly learn of the broken promise, meaning they get a clear, unbiased picture of what the promise was and that it was broken. This transfer is clear and neutral; it is a simple statement of the fact that the promise was broken. The citizens then decide that they disapprove of the action of the leader, and finally they decide to take action and punish the leader.

Figure 1: Basic Audience Cost Sequence



We focus on the link between Leaders and citizens: the transfer of information. Audience costs assume that citizens get an unbiased assessment of the leader’s broken promise; a neutral, true signal about the state of the world. While this is often employed as a simplifying assumption in the models, we might have doubts about whether this perfect neutral information transfer happens in the real world. If it does not, the effectiveness of audience costs can change: citizens might be more or less likely to punish their leaders, making “going public” a more or less effective signal.

## 2.2 Media, Framing, and Accountability

While audience cost theory itself is agnostic on how the information is transferred to the public, we assert that the media is important for information transmission, and that this transmission is rarely neutral. The media influences the opinion of the public through the framing of its content<sup>2</sup>. The media shapes how the public views events and influences public opinion through its framing of issues. The scholarly literature shows that frames can manipulate how consumers of media assign blame for policies, responsibilities for fixing them, their approval for social movements, etc. Clearly, the media plays a crucial role in both delivering information and shaping the way the public consumes and interprets it.

In this way, the media can also play a role in influencing audience costs. Scholars have already found evidence that the media, as well as other forms of information transmission like partisan rhetoric, can shape audience costs (Choi and James 2006; Potter and Baum 2010; Potter and Baum 2014; Slantchev 2006; Trager and Vavreck 2011). These authors make the claim that factors like media openness and a healthy political opposition group can make signals from leaders more believable, the logic being that open media allows for the unobstructed transfer of information and political opposition groups let opposing leaders know how likely the threatening leader is to follow up on their threat.

This interpretation, however, assumes that the media does not manipulate the message they are sending through the use of frames. It could be the case that when the media is reporting on the opposition, they are doing so in an unbiased matter. However, the framing literature suggests that this is not the case. What is more likely is that the media is taking the stories and wording them in certain ways. The way the media phrases these stories is going to influence how the domestic audience responds to them. Specifically, it could make

---

<sup>2</sup>See Chong and Druckman 2007a; Chong and Druckman 2007b; Chong and Druckman 2007c; Druckman 2001; Entman 2007; Iyengar 1990; Iyengar 1991; Kellstedt 2000; McLeod and Detenber 1999; Nelson, Oxley, and Clawson 1997; Scheufele 1999; Scheufele and Tewksbury 2006; Schuck and deVreese 2006; Simon and Xenos 2000; Slothuus and deVreese 2009.

the domestic audience more or less likely to punish a leader for renegeing on a threat. This makes the threat more or less useful as a signal of intent.

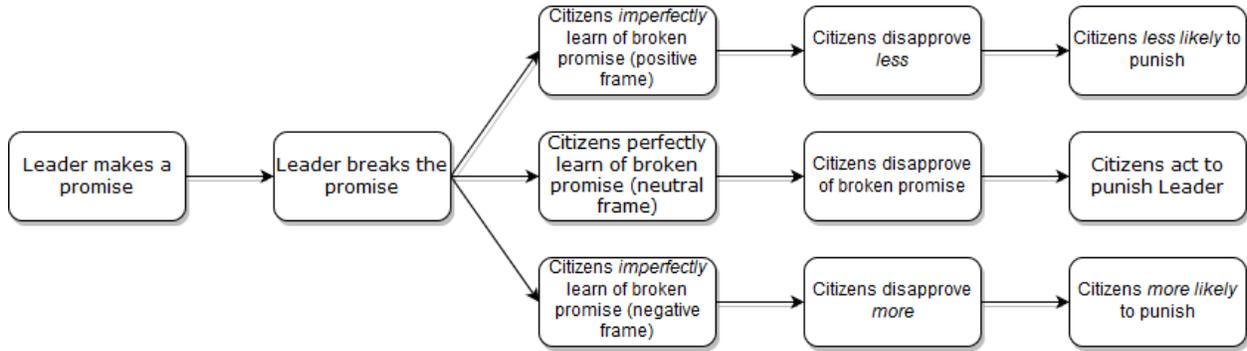
### 3 Theory

Audience cost theory, in summary, presents us with a situation where a leader makes some public claim. The domestic audience gains knowledge of this claim somehow and can credibly threatens to punish the leader if they renege on the promise made. This assumes perfect information on the part of the audience: they know the statement was made and subsequently broken. Often, this type of perfect knowledge is lacking in the real world, which leads to a situation where the public cannot credibly commit to punishing the leader. If this perfect, unbiased knowledge is lacking, the statement cannot be used as a signal of resolve: the opposing leader has no reason to believe the threatening leader if they believe they can do so without punishment from their domestic audience.

This suggests that the influence of the media is paramount when considering audience costs. The public receives their information about the leader's actions through the media, and the media frames the actions in a certain way. While there are a number of specific frames the media can use, they generally fall into two categories: positive and negative. These frames have differing implications for the likelihood of the leader being punished for backing down after issuing a threat which in turn leads to differences in how useful the initial escalation is as a signal of resolve. This can be seen outlined in the figure below.

As we can see from the figure, the addition of frames cause deviations from the basic logic of audience costs. Depending on how the news of the broken promise is framed, we see increased or decreased disapproval of the decision and consequently increased or decreased willingness to punish the leader, relative to the neutral (unbiased or "true") frame. This gives us two potential areas where the logic of audience costs can break: the level of public

Figure 2: Audience Cost Sequence with Imperfect Transfer



disapproval for the decision and the willingness to punish the executive.

A positive frame paints the decision to back down in a positive light. These stories might say that the decision was based on new information and that it was the correct course of action. They could also point out that the shift in policy saved lives, time, and money, clearly making it the better course of action. In effect, these frames are designed to mitigate the backlash the leader might have received otherwise. This results in *increased* approval (*decreased* disapproval); the Citizens are more accepting of the Leader’s decision to back down. As a result of this lower level of disapproval, Citizens are less likely to act to punish the Leader.

A negative frame, in contrast, is designed to highlight the negative aspects of breaking the promise. These frames suggest that the leader was weak-willed, caved to pressure or impugned the national honor in some other way. They might also present figures opposite of the positive frames, suggesting that lives may have been lost due to the leader’s threat of escalation and then backing down. These versions of events may make it seem like the Leader’s decision has put the Citizens’ way of life at risk. However they are worded, these frames are designed to generate negativity towards the Leader’s actions, resulting in *decreased* approval for the Leader’s decision and an *increased* probability of punishment, relative to the “neutral” frame.

In conclusion, there are two areas where framing makes a difference are the Citizens’

approval of the decision to break the promise and the Citizens' willingness to punish the Leader. Changes at either of these points break the logic of audience costs, drawing into question their usefulness as theoretical tools. This leads to two sets of hypotheses. First:

**Hypothesis 1:** When a positive (negative) frame is used, Citizens will express increased (decreased) support for the Leader's decision, relative to a neutral frame.

The next set of hypotheses deals with the actual decision to punish the Leader:

**Hypothesis 2:** When a positive (negative) frame is used, Citizens will be less (more) likely to punish the Leader.

## 4 Research Design

We test the theory using an experiment where subjects are drawn from the student population and randomly assigned to one of ten conditions. In the main set of conditions, respondents are given a vignette where they read about the President of the United States<sup>3</sup> threatening the use of military force against another country if that country does not cease a certain activity. The participants then read that the target state did not cease the activity and that the President backed down<sup>4</sup>. Respondents then are asked to fill out a survey with

---

<sup>3</sup>One might be concerned that using the President of the United States might influence our results. It could be that President Trump is so strongly disliked that our results may be affected. We do not consider this to be a problem for two reasons. First, the extreme dislike of the President suggests that audience costs should be really strong here; anything the President does should be subject to extreme scrutiny and punishment. This makes using the current President a "hard case" for our assertion that audience costs can be alleviated simply by using a positive frame. Second, our sample skews heavily liberal, as discussed below. Since this is the case, we expect everyone in the sample to have a relatively equal dislike of the President, meaning that any results we find are not due to differences of opinion regarding the current administration.

<sup>4</sup>"Backed down" implies that the President did not follow through on the threat and the status quo remains.

questions asking about their approval of the President’s actions as well as if they would be willing to vote for this President.

The main difference between these experimental conditions is in the frame put on the vignette<sup>5</sup>. The first condition, called the ‘No Frame’ or ‘Neutral’ condition, will have participants reading the facts of the situation without any attempt to manipulate them one way or the other. These vignettes will simply state that the President promised to take some action and then failed to do so. Previous work (Tomz 2007) states that we should expect to observe some form of audience cost in this condition; people tend to respond negatively to leaders reneging on their promises. Given this, the expectation is for there to be moderate to low approval of the decision and the President, and a low overall likelihood of voting for the President. We measure decision approval by asking respondents directly if they approve of the President’s decision to back down. Presidential approval is measured by asking respondents to rate approval for the President’s actions on a 0 to 100 scale. Finally, we look at willingness to directly punish the President by asking respondents if they would vote for the President in a future election. The results from the condition will then be used as a standard of comparison for the other two conditions.

The second condition adds a positive frame to the vignette. The information provided by the frame is unchanged, so these respondents still observe the President making and then breaking a promise. This frame adds a sentence praising the President for avoiding an unnecessary conflict. We expect a higher overall approval of the President and an increased willingness to vote for the President in this condition, in comparison to the baseline.

The final condition presents this information using a negative frame. Once again, the information presented in the vignette is exactly the same: the respondents learn that the

---

<sup>5</sup>A brief note on the vignettes themselves. The vignettes are written as statements without attribution, e.g., they are not meant to mimic an actual news story. We decided to write them this way mainly for the reason we outlined at the beginning of our discussion of frames: no source is going to be truly neutral, and attributing the statement to a source might carry connotations that would influence how respondents react to the vignettes.

President made a threat and then backed down. In this version of the vignette, however, it's noted that the President “flip-flopped” and it specifically points out that the President backed down when challenged. Thus, we expect respondents to express a lower approval of the President and a lower likelihood of voting for them relative to the baseline.

## 4.1 A Note About Frames

There is general agreement that there are two types of frames: emphasis and equivalence frames. These frames differ in their structure, but theoretically accomplish the same goal: they influence how people interpret information.

Equivalency frames are logically equivalent statements that cause people to have different reactions (Chong and Druckman 2007c; Druckman 2004; Tversky and Kahneman 1986). For example, the phrases “90% employment” and “10% unemployment” convey logically equivalent information, but they stress different things. The focus on employment versus unemployment, however, influences how people internalize this information and how they feel about it afterwards. In this example, respondents may report feeling more positive about the 90% employment statistic versus those who encounter the 10% statistic (Chong and Druckman 2007c).

Emphasis frames are what Druckman (2001) calls “frames in communication” (p. 227). These frames are designed to emphasize what the speaker sees as relevant to a particular topic. Druckman (2001) uses the example of a politician emphasizing the economy on the campaign trail, calling this the use of an “economic frame” (p. 227). The important point is that these types of frames draw attention to certain points over others, but they do not necessarily be about the same topic. These frames can be comparing completely different things; what is of importance here is that they prompt different emotional responses to the information.

We use emphasis frames in the research design laid out above. The President making

a decision based on new information and the President being a coward are not logically equivalent; they are only intended to trigger different emotional responses. We do not have a theoretical reason to believe that changing from an emphasis frame to an equivalency frame will have an effect, but the design can be changed to incorporate them to test this assertion. Additionally, the literature considers them to be two distinct types of frames so it is important to test both types. In these conditions, the positive frame states that the President’s actions protected the safety of 80% of the people at risk. The negative frame, on the other hand, states that the President’s actions endangered the safety of 20% of the people at risk. Note that these frames are logically equivalent: the same percentage of people were protected/endangered in both. The difference lies in how this information is presented to the respondent, with the positive frame focusing on lives saved and the negative frame focusing on lives endangered. Once again, we expect that in the positive condition people should be more likely to support the President and the decision to back down, and they should be less likely to support the President and the decision to back down after encountering the negative condition.

## **4.2 A Note About Context**

One final thing to be potentially concerned about is the nature of the broken promise itself. In the main set of vignettes, we follow the example of Tomz (2007) and lay out a situation where the threatened action was military retaliation. This is the standard application for audience costs: they are used to explain situations where a threat of retaliation caused both sides to avoid some military conflict. Audience costs need not be limited to these contexts, and we might expect respondents to have more intense reactions to these scenarios. As Kertzer and Brutger (2016) point out, a leader can incur audience costs simply by threatening the use of force, and we need to be able to separate “true” audience costs from “belligerence” costs. As a way to account for this, we included a full set of vignettes

that described a situation where the President threatened *sanctions* and then backed down from implementing them. Since sanctions are generally considered to be an alternative to the use of force, we believe that this should reduce any sort of “belligerence” costs.

The first set of vignettes uses the same positive, neutral, and negative frame set-up outlined above. The neutral frame vignette simply states that the country did not stop when threatened with the sanctions and the President did not follow through with them. The positive frame vignette again says that the President did not follow through with the sanctions because he learned new information. Finally, the negative frame vignette again attributes his decision to back down to flip-flopping.

The second set mimics the equivalency frames outlined above. In the positive equivalency frame, the vignette states that the President’s decision to break his promise protected the financial well-being of about 80% of the people at risk. The negative equivalency frame states that the President’s decision endangered the financial well-being of 20% of the people at risk. Again, the number of people protected and put at risk by the President’s decision is the same in both of these conditions; the main difference between them is what is emphasized by the frame.

## 5 Results

Our data comes from a sample of 1020 students currently enrolled at the University of California, Merced. 29% of the sample is male and 70% is female, with the remainder either failing to answer or selecting the “Other” option. Ages range from 17 to 39 years old, with the modal age being 18. 84% of respondents indicated they at least leaned liberal, while 15% indicated that they at least leaned conservative<sup>6</sup>.

---

<sup>6</sup>The distribution was similar for party identification.

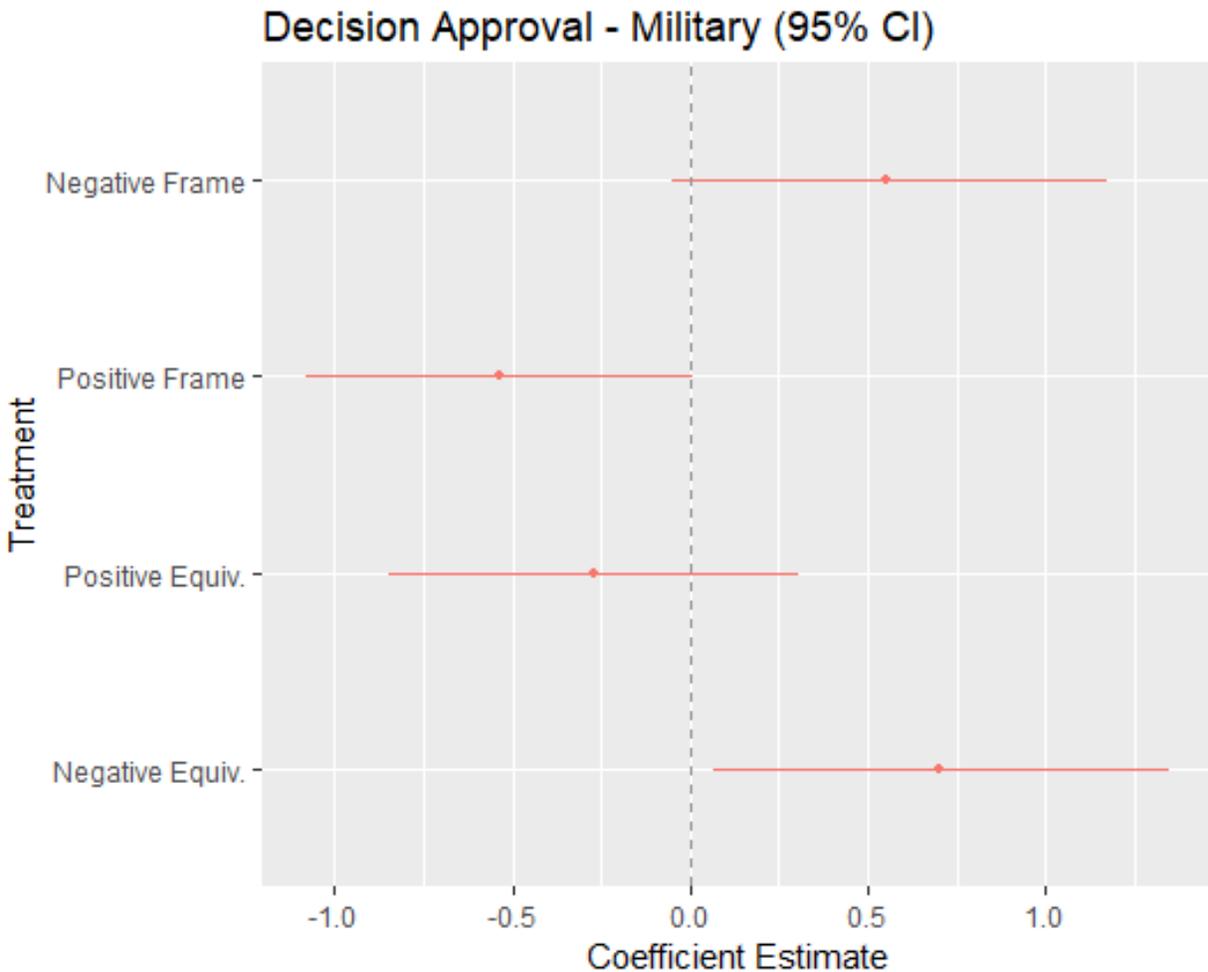
## 5.1 Hypothesis 1

Our first logit model looks at the effect of frame on approval of the President's decision to back down for the military vignettes. The results are reported in the table below. The "decision approval" variable is binary, with a score of one denoting approval of the decision to back down and a score of 2 denoting disapproval of the decision. Thus, we expect that the negative frames should have a *positive* coefficient, indicating increased disapproval, while the positive frames should have a *negative* coefficient. Recall that these are in relation to the neutral frame, so a statistically significant response indicates that the frame is different from the effect the neutral frame had on decision approval.

Table 1: Effect of Frame on Decision Disapproval (Military)

	<i>Dependent variable:</i>
	Decision Disapproval, Military
Negative Frame	0.555* (0.313)
Positive Frame	-0.534* (0.277)
Positive Equivalency	-0.270 (0.294)
Negative Equivalency	0.699** (0.326)
Constant	0.649*** (0.209)
Observations	520
Log Likelihood	-319.649
Akaike Inf. Crit.	649.298
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

Figure 3: Effect of Frame on Decision Disapproval (Military)



As Table 1 demonstrates, three of our military frames had a significant effect (at least the  $p < 0.1$  level) on decision approval: the Negative frame, the Positive frame, and the Negative Equivalency Frame. These frames had the predicted effect on the respondent's approval of the President's decision: those who received the negative frames were *more likely* to disapprove, while those who got the positive frame were *less likely* to disapprove<sup>7</sup>, relative to the neutral frame. We do, however, want to ensure our results are not subject to belligerence costs (Kertzer and Brutger 2016). If this were the case, it might artificially inflate the negative

<sup>7</sup>While the Positive Equivalency frame was not statistically significant, the coefficient suggests it acted as predicted.

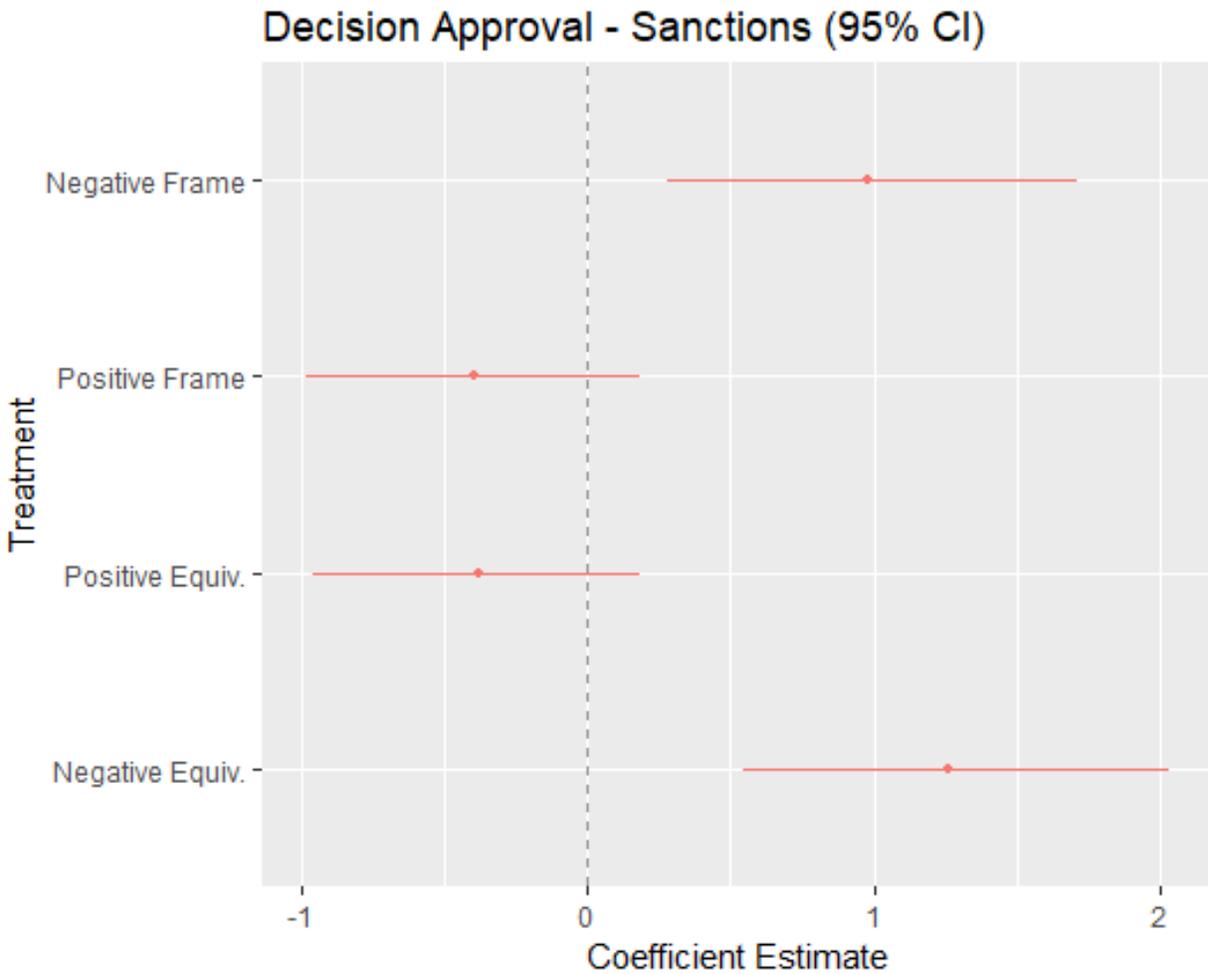
results, as respondents act more negatively towards a president suggesting military action. We might also expect these belligerence costs to cancel out the effects of a positive frame: the respondent might have positive feelings about the decision due to the frame, but these might be counter-acted by the negative feelings from the threat of force. To further examine this, our next logit model looks at how frames influence decision approval in the sanctions context. The results are reported in Table 2.

We see a similar pattern: the positive frames decrease the likelihood of disapproving of the action (negative coefficients) and negative frames increase the likelihood of disapproving of the President's action (positive coefficient). Both of the negative frames are positive and significant (at the  $p < 0.05$  level), and while neither of the positive frames are significant, their coefficients are in the expected direction. This might be because of the nature of sanctions: it might be more difficult to put a similar positive "spin" on failing to sanction because they are so ephemeral. Saying the President avoided military conflict thanks to new information can easily translate into the President saving lives, while the effect of avoiding sanctions is not so easily extrapolated. Given the results across both of these models, we conclude that there is mixed support for Hypothesis 1: depending on the context, positive frames decrease disapproval with a Leader's decision to renege, while negative frames increase disapproval with the Leader's decision to renege.

Table 2: Effect of Frame on Decision Disapproval (Sanctions)

	<i>Dependent variable:</i>
	Decision Disapproval, Sanctions
Negative Frame	0.977*** (0.363)
Positive Frame	-0.395 (0.297)
Positive Equivalency Frame	-0.381 (0.291)
Negative Equivalency Frame	1.263*** (0.377)
Constant	0.647*** (0.215)
Observations	476
Log Likelihood	-273.902
Akaike Inf. Crit.	557.805
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

Figure 4: Effect of Frame on Decision Disapproval (Sanctions)



## 5.2 Hypothesis 2

### 5.2.1 Presidential Approval

Hypothesis 2, meanwhile, looks at whether Citizens can channel this disapproval into action. This hypothesis states that those who got the information through a positive frame will be less likely to act to punish the President than those who got a neutral frame, and those who got a negative frame will be more likely to act to punish the President than those who got a neutral frame. As a hard test, we look at those who already disapprove of the President's decision: if they already disapprove, a negative frame should not have an effect

on their willingness to act due to floor effects, and a positive frame should not be enough to make them support a leader who's decisions they disapprove of. We look at willingness to punish two ways: first using presidential approval with the expectation that lower approval means a lower likelihood of support in the next election, and second using the respondent's stated willingness to vote for the President in the next election. The first step is to limit our sample to those who disapproved of the President's decision in both the sanction and military context. This leaves us with 667 observations; 340 for the military context and 327 for the sanction context. Next, we conduct an ANOVA for the effect of treatment on the military sample; the results are reported in Table 3 (the results for the full sample, including those who approve of the President's decision, can be found in Table 4).

Table 3: ANOVA for Presidential Approval (Military)

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Treatment Group	4	2708.49	677.12	1.45	0.2183
Residuals	321	150234.73	468.02		

Figure 5: Presidential Approval Across Treatments (Military)

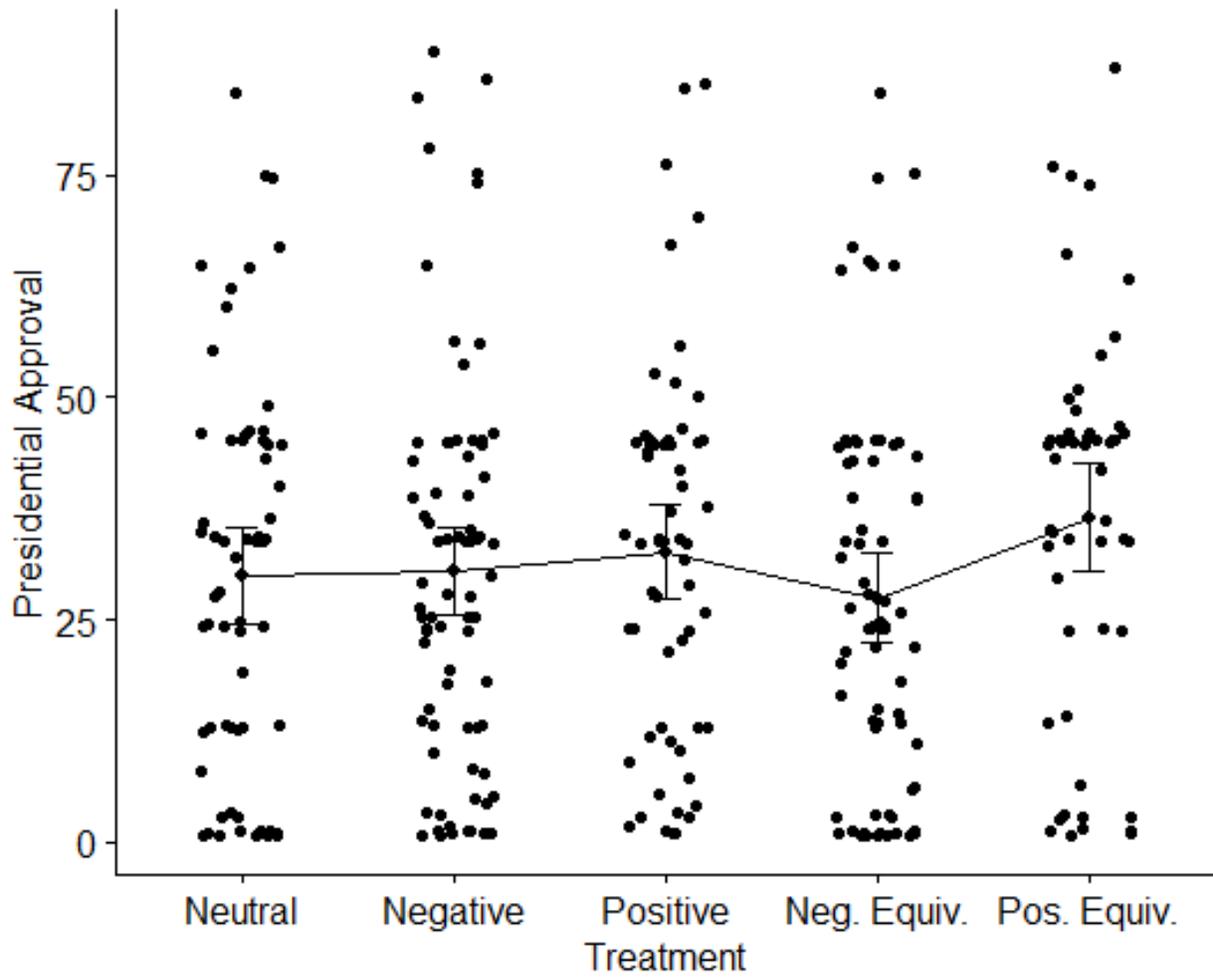
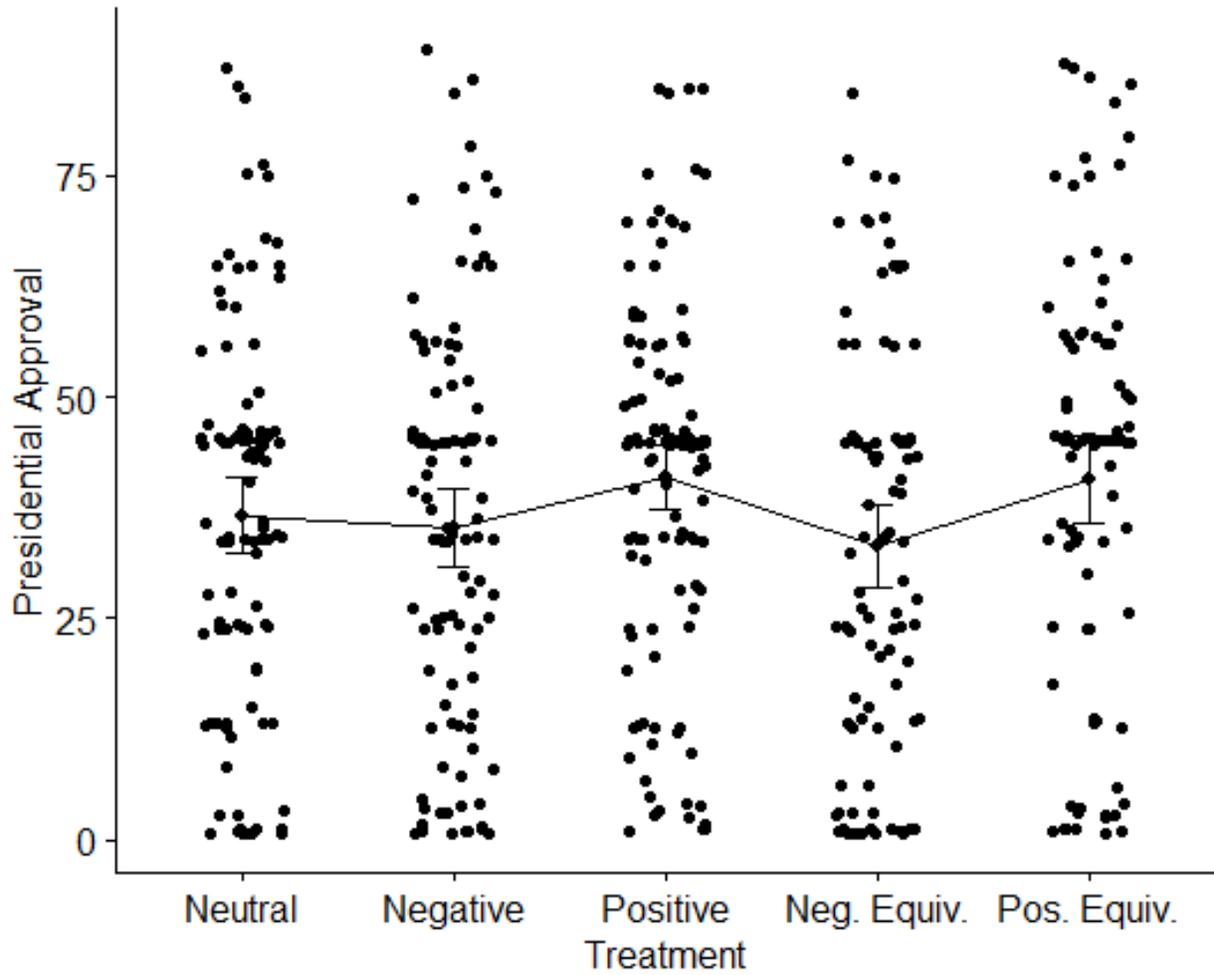


Table 4: ANOVA for Presidential Approval (Military - Full Sample)

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
treatment_mil	4	4724.26	1181.06	2.38	0.0510
Residuals	508	252387.46	496.83		

Figure 6: Presidential Approval Across Treatments (Military - Full Sample)



This table tells us that, for the military context, there is no difference in presidential approval between the treatment groups for respondents who disapprove of the President’s decision; Figure 5 shows this more explicitly. As we can see, the mean level of presidential approval is relatively stable across the different treatment conditions, suggesting that the frame does not matter for presidential approval in the military complex. It does appear that the equivalency frames have some minor effect, decreasing approval for the negative frame and increasing it for the positive, but this difference is not significant.

Next, we look at presidential approval in the sanctions context; the findings for this analysis are presented in Table 5 (the full sample is presented in Table 6). Again, we see no significant difference between groups in the disapproval sample; Figure 7 shows this more explicitly.

Table 5: ANOVA for Presidential Approval (Sanctions)

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Treatment Group	4	2740.16	685.04	1.58	0.1786
Residuals	307	132842.97	432.71		

Figure 7: Presidential Approval Across Treatments (Sanctions)

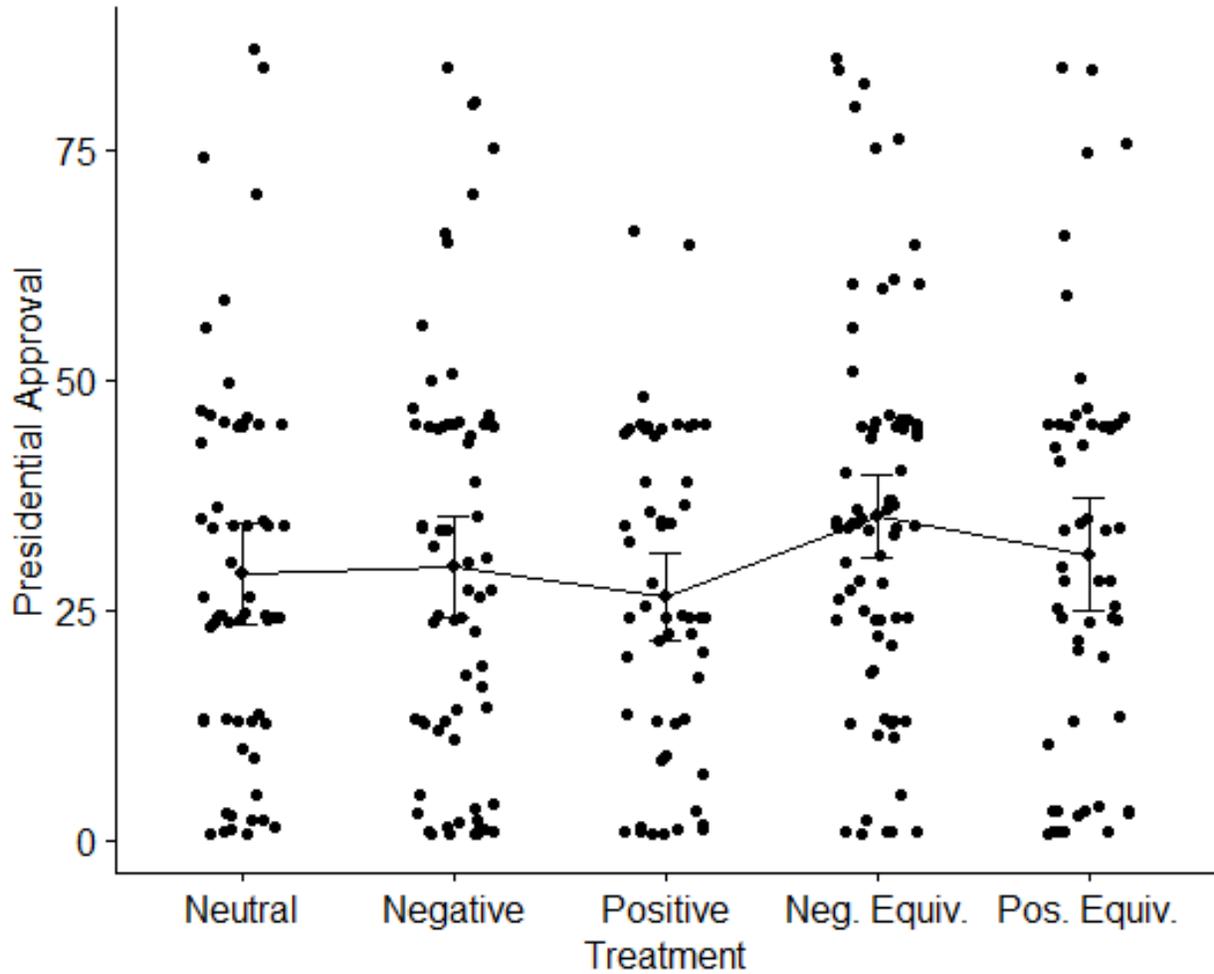
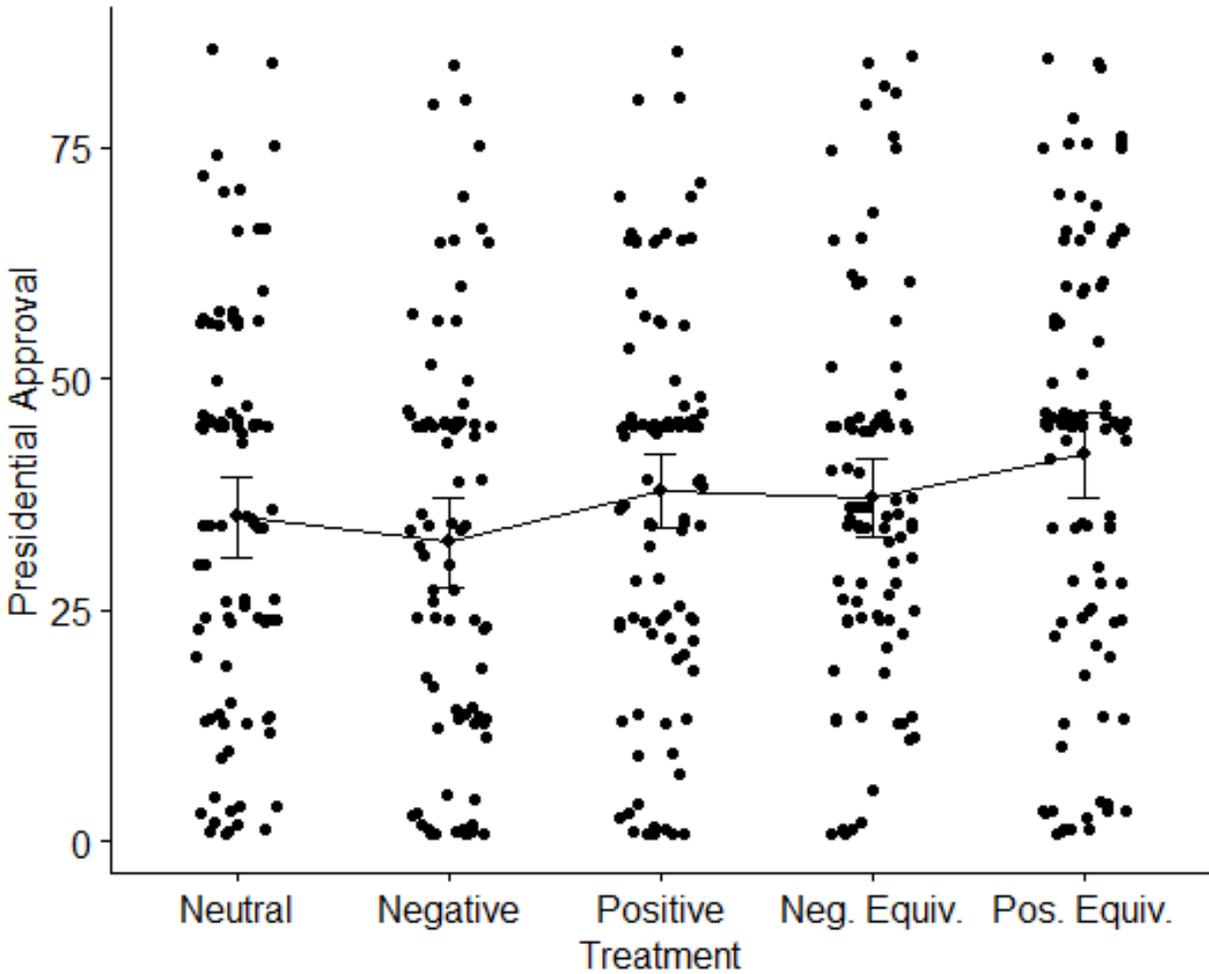


Table 6: ANOVA for Presidential Approval (Sanctions - Full Sample)

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
treatment_sanc	4	4471.66	1117.92	2.42	0.0480
Residuals	465	215089.49	462.56		

Figure 8: Presidential Approval Across Treatments (Sanctions - Full Sample)



### 5.2.2 A Note on Vote Choice

Presidential approval is, admittedly, an imperfect measure of individual's actions to punish the president. We are inferring that a lower/higher presidential approval translates into a higher/lower probability of punishment. In an attempt to get around this issue, we also asked respondents if they would be willing to vote for the President in an upcoming election. We were specifically interested in those who expressed disapproval of the President's decision. Our theory suggests that audience costs can break down in two places, disapproval and action, and it implies that they are separate from one another. Thus, we were interested

in seeing if those who disapproved of the President’s action would still be willing to vote for them if they received the positive frame. Unfortunately, the number of respondents who reported a willingness to vote for the President after they disapproved of his actions was so low (15 in the military context, 14 in the sanctions context), that would could not draw any conclusions empirically. The results suggest, however, that once someone disapproves of the President’s action, they are willing to punish them<sup>8</sup>.

## 6 Implications & Conclusions

In this paper, we outlined the potential ways that frames could influence the logic of audience costs. The existing literature says that leaders can make credible public declarations because other leaders know that the domestic public can be counted on to punish a leader that backs down from a public declaration. It is assumed that if a leader backs down, the public will learn about the broken promise through some neutral and perfectly informative method, they will disapprove of this decision, and will proceed to punish the leader.

We demonstrate that this may not be the case. People often learn about political news through the media, which is never neutral and rarely perfectly accurate. These differences matter because they can lead to the citizens getting different versions of events, some framing the decision in a positive light and others giving the decision a negative spin.

Our results show that the way a story is framed has a statistically significant effect on whether citizens support the leader’s decision to break their promise. In the military context, positive frames make citizens less likely to disapprove of the President’s actions while negative frames make them more likely. When respondents are told that the broken promise involved sanctions, negative frames again make them more likely to disapprove. What is less clear is

---

<sup>8</sup>Another possibility is simply that “President” is currently a loaded term. We discussed above why this should not be a problem generally, but it could be an issue here: given how liberal our sample is, there might not be any variation because liberal approval of President Trump is so low.

whether the translation of disapproval into political action is influenced by frames.

While we do not find complete support for our theory, the results we do find have serious implications for audience costs. Our results show that citizens will be more or less willing to support their leader's decision to renege on promises depending on how news of the broken promise is framed. If it is framed negatively, then the basic logic of audience costs holds: citizens will disapprove of backing down, which means that a public declaration is a credible signal. If, on the other, the leader that makes a public declaration has a friendly media environment, the story becomes a lot more complicated. News of a broken promise will be delivered in a more positive light, meaning that less of the public will disagree with the decision. This makes it unclear if the public commitment by the leader is credible, which can have consequences up to and including war.

Future work can take these results further in two ways. First, it can attempt to gather a larger, more representative sample of the population in attempt to replicate our results. We are working on this step ourselves, hoping that we can solve our issues by employing simulations. Second, future research should investigate the elite side of the equation. Does it matter that the public will be willing to support a decision to renege on a promise? Do elites take the domestic public into account when considering a public declaration by another leader? The answers to these questions will be extremely useful in determining when and how audience costs can be employed as a credible signal of intent and, hopefully, a deterrent for war.

## References

- Baum, Matthew A. (2004). “Going Private: Public Opinion, Presidential Rhetoric, and the Domestic Politics of Audience Costs in U.S. Foreign Policy Crises”. In: *Journal of Conflict Resolution* 48.5, pp. 603–631.
- Chaudoin, Stephen (2014). “Audience Features and the Strategic Timing of Trade Disputes”. In: *International Organization* 68, pp. 877–911.
- Choi, Seung-Whan and Patrick James (2006). “Media Openness, Democracy and Militarized Interstate Disputes”. In: *British Journal of Political Science* 37, pp. 23–46.
- Chong, Dennis and James N. Druckman (2007a). “A Theory of Framing and Opinion Formation in Competitive Elite Environments”. In: *Journal of Communication* 57, pp. 99–118.
- (2007b). “Framing Public Opinion in Competitive Democracies”. In: *American Political Science Review* 101.4, pp. 637–655.
- (2007c). “Framing Theory”. In: *Annual Review of Political Science* 10, pp. 103–126.
- Conrad, Courtenay R., Justin Conrad, and Joseph K. Young (2014). “Tyrants and Terrorism: Why Some Autocrats are Terrorized While Others are Not”. In: *International Studies Quarterly* 28, pp. 539–549.
- Davies, Graeme A. M. and Robert Johns (2013). “Audience Costs among the British Public: The Impact of Escalation, Crisis Type, and Prime Ministerial Rhetoric”. In: *International Studies Quarterly* 57, pp. 725–737.
- Downes, Alexander B. and Todd S. Sechser (2012). “The Illusion of Democratic Credibility”. In: *International Organization* 66, pp. 457–489.
- Druckman, James N. (2001). “The Implications of Framing Effects for Citizen Competence”. In: *Political Behavior* 23.3, pp. 225–256.

- Druckman, James N. (2004). "Political Preference Formation: Competition, Deliberation, and the (Ir)relevance of Framing Effects". In: *American Political Science Review* 98.4, pp. 671–686.
- Entman, Robert M. (2007). "Framing Bias: Media in the Distribution of Power". In: *Journal of Communication* 57.1, pp. 163–173.
- Eyerman, Joe and Robert A. Hart (1996). "An Empirical Test of the Audience Cost Proposition: Democracy Speaks Louder than Words". In: *Journal of Conflict Resolution* 40.4, pp. 597–616.
- Fearon, James D. (1994a). "Domestic Political Audiences and the Escalation of International Disputes". In: *American Political Science Review* 88.3, pp. 577–592.
- (1994b). "Signaling versus the Balance of Power and Interests: An Empirical Test of a Crisis Bargaining Model". In: *Journal of Conflict Resolution* 38.2, pp. 236–269.
- Gibler, Douglas M. and Marc L. Hutchison (2013). "Territorial Issues, Audience Costs, and the Democratic Peace: The Importance of Issue Salience". In: *Journal of Politics* 75.4, pp. 879–893.
- Guisinger, Alexandra and Alastair Smith (2002). "Honest Threats: The Interaction of Reputation and Political Institutions in International Crises". In: *Journal of Conflict Resolution* 46.2, pp. 175–200.
- Hoffman, Aaron M. et al. (2009). "Satisfaction, Alternatives, Investments, and the Microfoundations of Audience Cost Models". In: *International Interactions* 35.4, pp. 365–389.
- Iyengar, Shanto (1990). "Framing Responsibility for Political Issues: The Case of Poverty". In: *Political Behavior* 12.1, pp. 19–40.
- (1991). *Is Anyone Responsible? How Television Frames Political Issues*. University of Chicago Press.
- Kellstedt, Paul M. (2000). "Media Framing and the Dynamics of Racial Policy Preferences". In: *American Journal of Political Science* 44.2, pp. 245–260.

- Kertzer, Joshua D. and Ryan Brutger (2016). “Decomposing Audience Costs: Bringing the Audience Back into Audience Cost Theory”. In: *American Journal of Political Science* 60.1, pp. 234–249.
- Kinne, Brandon J. and Nikolay Marinov (2012). “Electoral Authoritarianism and Credible Signaling in International Crises”. In: *Journal of Conflict Resolution* 57.3, pp. 359–386.
- Kurizaki, Shuhei and Taehee Whang (2015). “Detecting Audience Costs in International Disputes”. In: *International Organization* 69, pp. 949–980.
- Levendusky, Matthew S. and Michael C. Horowitz (2012). “When Backing Down is the Right Decision: Partisanship, New Information, and Audience Costs”. In: *Journal of Politics* 74.2, pp. 323–338.
- Levy, Jack S. et al. (2015). “Backing Out or Backing In? Commitment and Consistency in Audience Costs Theory”. In: *American Journal of Political Science* 59.4, pp. 988–1001.
- McLeod, Douglas M. and Benjamin H. Detenber (1999). “Framing Effects of Television News Coverage of Social Protest”. In: *Journal of Communication* 49.3, pp. 3–23.
- McManus, Roseanne W. (2014). “Fighting Words: The Effectiveness of Statements of Resolve in International Conflict”. In: *Journal of Peace Research* 51.6, pp. 726–740.
- Nelson, Thomas E., Zoe M. Oxley, and Rosalee A. Clawson (1997). “Toward a Psychology of Framing Effects”. In: *Political Behavior* 19.3, pp. 221–246.
- Partell, Peter J. and Glenn Palmer (1999). “Audience Costs and Interstate Crises: An Empirical Assessment of Fearon’s Model of Dispute Outcomes”. In: 43, pp. 389–405.
- Potter, Philip B. K. and Matthew A. Baum (2010). “Democratic Peace, Domestic Audience Costs, and Political Communication”. In: *Political Communication* 27.4, pp. 453–470.
- (2014). “Looking for Audience Costs in all the Wrong Places: Electoral Institutions, Media Access, and Democratic Constraint”. In: *Journal of Politics* 76.1, pp. 167–181.

- Prins, Brandon C. (2003). “Institutional Instability and the Credibility of Audience Costs: Political Participation and Interstate Crisis Bargaining, 1816–1992”. In: *Journal of Peace Research* 40.1, pp. 67–84.
- Ramsay, Kristopher W. (2004). “Politics at the Water’s Edge: Crisis Bargaining and Electoral Competition”. In: *Journal of Conflict Resolution* 48.4, pp. 459–486.
- Scheufele, Dietram A. (1999). “Framing as a Theory of Media Effects”. In: *Journal of Communication* 49.1, pp. 103–122.
- Scheufele, Dietram A. and David Tewksbury (2006). “Framing, Agenda Setting, and Priming: The Evolution of Three Media Effects Models”. In: *Journal of Communication* 57 (1), pp. 9–20.
- Schuck, Andreas R. T. and Claes H. deVreese (2006). “Between Risk and Opportunity: News Framing and its Effects on Public Support for EU Enlargement”. In: *European Journal of Communication* 21.1, pp. 5–32.
- Schultz, Kenneth A. (2001). “Looking for Audience Costs”. In: *Journal of Conflict Resolution* 45.1, pp. 32–60.
- Simon, Adam and Michael Xenos (2000). “Media Framing and Effective Public Deliberation”. In: *Political Communication* 17.4, pp. 363–376.
- Slantchev, Branislav L. (2006). “Politicians, the Media, and Domestic Audience Costs”. In: *International Studies Quarterly* 50, pp. 445–477.
- (2012). “Audience Cost Theory and Its Audiences”. In: *Security Studies* 21.3, pp. 376–382.
- Slothuus, Rune and Claes H. deVreese (2009). “Political Parties, Motivated Reasoning, and Issue Framing Effects”. In: *Journal of Politics* 72.3, pp. 630–645.
- Smith, Alastair (1998). “International Crises and Domestic Politics”. In: *American political Science Review* 92.3, pp. 623–638.

- Snyder, Jack and Erica D. Borghard (2011). “The Cost of Empty Threats: A Penny, Not a Pound”. In: *American Political Science Review* 105.3, pp. 437–456.
- Tarar, Ahmer and Bahar Leventoğlu (2009). “Public Commitment in Crisis Bargaining”. In: *International Studies Quarterly* 53, pp. 817–839.
- (2012). “Limited Audience Costs in International Crises”. In: *Journal of Conflict Resolution* 57.6, pp. 1065–1089.
- Tingley, Dustin H. and Barbara F. Walter (2011). “Can Cheap Talk Deter? An Experimental Analysis”. In: *Journal of Conflict Resolution* 55.6, pp. 996–1020.
- Tomz, Michael (2007). “Domestic Audience Costs in International Relations: An Experimental Approach”. In: *International Organization* 61, pp. 821–840.
- Trachtenberg, Marc (2012). “Audience Costs: An Historical Analysis”. In: *Security Studies* 21.1, pp. 3–42.
- Trager, Robert F. and Lynn Vavreck (2011). “The Political Costs of Crisis Bargaining: Presidential Rhetoric and the Role of Party”. In: *American Journal of Political Science* 55.3, pp. 526–545.
- Tversky, Amos and Daniel Kahneman (1986). “Rational Choice and the Framing of Decisions”. In: *Journal of Business* 59.4, S251–S278.
- Uzonyi, Gary, Mark Souva, and Sona N. Golder (2012). “Domestic Institutions and Credible Signals”. In: *International Studies Quarterly* 56, pp. 765–776.
- Weeks, Jessica L. (2008). “Autocratic Audience Costs: Regime Type and Signaling Resolve”. In: *International Organization* 62, pp. 35–64.

## A Survey Questions

### A.1 Emphasis Frames

Positive Military Emphasis Frame:

The President of the United States previously threatened military force against another country if the country does not cease an action the United States deems inappropriate. The country did not stop but the president did not follow through because he learned new information, no military force was taken against them.

Positive Sanction Emphasis Frame:

The President of the United States previously threatened sanctions against another country if the country does not cease an action the United States deems inappropriate. The country did not stop but the president did not follow through because he learned new information, no sanctions were made against them.

Negative Military Emphasis Frame:

The President of the United States previously threatened military force against another country if the country does not cease an action the United States deems inappropriate. The country did not stop but the president did not follow through because he flip-flopped, no military force was taken against them.

Negative Sanction Emphasis Frame:

The President of the United States previously threatened sanctions against another country if the country does not cease an action the United States deems inappropriate. The country did not stop but the president did not follow through because he flip-flopped, no sanctions were made against them.

Neutral Military Emphasis Frame:

The President of the United States previously threatened military force against another country if the country does not cease an action the United States deems inappropriate. The country did not stop but the president did not follow through, no military force was taken against them.

Neutral Sanction Emphasis Frame:

The President of the United States previously threatened sanctions against another country if the country does not cease an action the United States deems inappropriate. The country did not stop but the president did not follow through, no sanctions were made against them.

## **A.2 Equivalency Frames**

Positive Military Equivalency Frame:

The President of the United States previously threatened military force against another country if the country does not cease an action the United States deems inappropriate. The country did not stop but the president did not follow through, no military force was taken against them. This decision protected the safety of about 80% of the people at risk.

Positive Sanction Equivalency Frame:

The President of the United States previously threatened sanctions against another country if the country does not cease an action the United States deems inappropriate. The country did not stop but the president did not follow through, no sanctions were made against them. This decision protected the financial well-being of about 80% of the people at risk.

Negative Military Equivalency Frame:

The President of the United States previously threatened military force against another country if the country does not cease an action the United States deems inappropriate. The country did not stop but the president did not follow through, no military force was taken against them. This decision endangered the safety of about 20% of the people at risk.

Negative Sanction Equivalency Frame:

The President of the United States previously threatened sanctions against another country if the country does not cease an action the United States deems inappropriate. The country did not stop but the president did not follow through, no sanctions were made against them. This decision endangered the financial well-being of about 20% of the people at risk.