Backing Up, Not Backing Down:
Mitigating Audience Costs Through Policy Substitution

(10,349 words, excluding title page)

1 August 2017

[UNDER REVIEW: Please do not cite without permission of author.]

Erik Lin-Greenberg[1]
Department of Political Science
Columbia University

ABSTRACT

Can a leader reduce the audience costs imposed for backing down completely on a threat by opting instead to "back up" to a less hawkish policy? Current research examines the political repercussions of making a threat and then taking no action at all. Real world officials, however, often "back up" and implement policies that involve some sort of action – for instance, imposing sanctions after threatening military force. Existing research overlooks whether this policy substitution reduces audience costs and, in turn, degrades the credibility of crisis signaling. This paper puts forth a theory of audience cost mitigation through policy substitution: backing up to less hawkish policies – that reduce inconsistency between a leader's words and deeds – may reduce audience costs. An original survey experiment fielded on 1,006 respondents finds support for the theory. The findings suggest audience costs are controllable, challenging the notion that leaders rarely bluff during crises.

Key Words: Audience Costs, Policy Substitution, Credibility, Crisis Bargaining, Reputation, Survey Experiment

[1] I am grateful to Richard Betts, Brian Blakenship, Keith Christopher, Alex de la Paz, Kolby Hanson, Robert Jervis, Kevin McKee, Roseanne McManus, Sara Moller, Theo Milonopoulos, Constantino Pischedda, Tonya Putnam, Tara Slough, Jordan Sorensen, Jack Snyder, and participants at the 2017 International Studies Association Conference and the Columbia Security Studies Working Group for their valuable comments and assistance. All errors are my own.

In 2012, President Obama publicly issued a red line against Syria's Assad regime, threatening military action if the government used chemical weapons. When Syrian forces employed chemical agents in 2013, the Obama administration backed up to a peaceful United Nations-mediated solution rather than fulfilling its threat of armed reaction. Critics chastised the administration for harming America's reputation, but Obama supporters argued that the President had achieved his policy objectives. Can state leaders mitigate the audience costs they incur for backing down from a threat by turning to less hawkish actions, like those taken by the Obama administration in 2013?

A vast body of research in international relations suggests that governments can credibly signal their intentions during crisis bargaining by generating domestic audience costs. According to audience cost theory, leaders tie their hands by publicly threatening to use force against a potential adversary. Since domestic constituents can punish their leaders for being inconsistent, leaders have incentives to avoid bluffing, which in turn makes their threats more credible (Fearon 1994; Schultz 1998; Weeks 2008). Existing empirical studies on audience costs, however, have unrealistically simplified the inconsistency between threats and actions, typically treating whether a leader follows through on threats as a dichotomous variable (Tomz 2007; Levendusky and Horowitz 2012; Kertzer and Brutger 2016). In other words, states in these studies either carry out their threat completely or *"back down"* by taking no action at all. This simplifying assumption is problematic. In practice, there can be significant variation in the degree of inconsistency between a leader's words and deeds. Some leaders might follow through with their threats, while others can *"back up"* to a range of less hawkish actions after initially making a threat to use force. If leaders face less punishment for backing up than they do for

backing down, they may be more apt to bluff about their willingness to use force, degrading the credibility of crisis signaling.

This paper introduces and tests a theory that suggests leaders can minimize the perceived inconsistency of their words and deeds – and the associated audience costs – by backing up, which I define as making a threat before adopting some less hawkish action instead of doing nothing at all. Does backing up instead of backing down allow leaders to mitigate audience costs? If so, why does the public deal less harshly with leaders who back up instead of back down? Most importantly, what does this mean for the credibility of signaling during crises? If backing up allows a leader to reduce audience costs, leaders may make strong threats before subsequently backing up to less forceful options. The binary treatment of following-through or backing-down in existing audience cost studies, therefore, overlooks an important dynamic that potentially shapes the credibility of signals. Unlike traditional audience cost experiments which characterize a leader as either a truthful type or a complete bluffing type, a state's ability to back up to a variety of policy options means that there may be a range of bluffing types. This variation may make it more difficult for adversaries to distinguish credible threats from bluffs, and lead states to take escalatory moves if they believe a rival may eventually back up from an initial threat.

To assess the costs and implications of backing up, I first synthesize the disparate literatures on audience costs, consistency, and policy substitution to develop a theory on the domestic political repercussions of partially reneging on threats. I then test this theory using an original survey experiment fielded on 1,006 respondents in the United States. The experiment exposes respondents to a hypothetical crisis scenario in which an allied state is invaded. I first manipulate whether the President threatens to deploy ground forces to defend the ally, and then

vary whether the President follows through on his threat, backs up to airstrikes or sanctions, or backs down altogether. I find that by backing up rather than backing down, leaders can significantly reduce the audience costs they suffer. Micro-level analysis of respondent justifications suggests that much of this reduction can be attributed to the perception that taking some form of action is more consistent behavior than reneging entirely. This analysis of microfoundations sheds light on how the public links variation in the *degree* of inconsistency between a President's threats and actions with perceptions of reputational consequences and Presidential competence, something that is largely absent from existing research. The finding that leaders can use policy substitution to reduce punishment for reneging from threats suggests leaders may have higher incentives to bluff than traditional audience cost theory predicts.

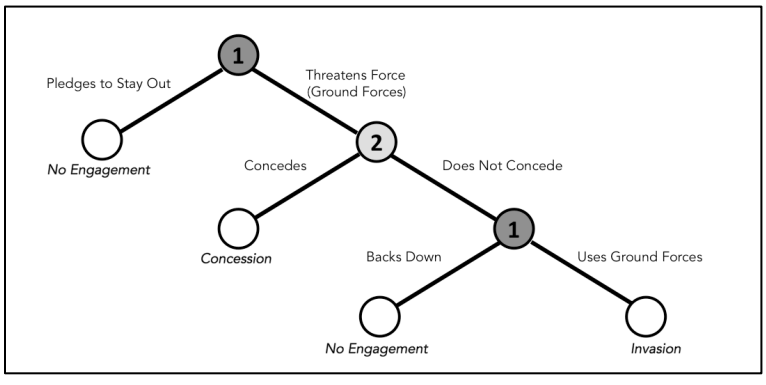
MITIGATING AUDIENCE COSTS THROUGH POLICY SUBSTITUTION

Audience Cost Theory

A large body of literature suggests that domestic audiences in both democratic and autocratic states punish their leaders for making a foreign policy threat and then reneging on it. backing down on a threat weakens a state's reputation by discrediting the value of its threats and promises and highlights the incompetence of a state's leader (Fearon 1994; Schultz 1998; Weeks 2008; Guisinger and Smith 2002). The high political costs for bluffing make leaders reluctant to issue threats with which they do not intend to follow through, enhancing the credibility of threats made by leaders of states that can generate significant audience costs (Fearon 1994). In other words, the potential for large political repercussions helps differentiate costly (and credible) signals from less credible cheap talk (Jervis 1970; Fearon 1994), and allow leaders to transmit information about their willingness to fight.

Audience cost theories are grounded in a classic crisis bargaining model that features two players: a challenger state (depicted as the dark grey '*State 1*' in the stylized game diagram in Figure 1) and its foreign adversary (the light grey '*State 2*'). At stake is some good (i.e. territory, a political objective), which for simplicity I assume is indivisible and has a normalized value of 1 in the eyes of each state's leader. The state that possesses this good receives a payoff of 1, while the state that does not receives a payoff of 0. At the start of the crisis, State 1 (the challenger) must decide whether to issue a threat of force to obtain the good. If the challenger pledges to stay out and does not issue a threat, there is no engagement and the status quo is maintained with State 2 retaining control of the good. If, however, State 1 threatens to use force, State 2 must then decide whether to concede. If State 2 concedes, it transfers control of the good to State 1. If State 2 does not concede, State 1 must then decide whether to follow through with its initial threat of force or back down and do nothing. If State 1 backs down, it does not gain control of the good and domestic audiences in State 1 punish their leader for reneging on her initial threat. If State 1 follows through and uses force, both states incur the costs of war and each has some probability of gaining control of the good.[2]

FIGURE 1: Classic Crisis Bargaining Model



[2] For a detailed examination of crisis bargaining models see Bruce Bueno de Mesquita. 1983. *The War Trap*. New Haven: Yale University Press.

The relationship between audience costs and the credibility of signals during crisis bargaining is grounded in the assumption that leaders are unable to control the political repercussion for backing down on a threat. However, recent scholarship has examined factors that shape a leader's vulnerability to audience costs, revealing various mechanisms that allow leaders to mitigate the consequences for not following through on their threats. Audience costs may be shaped by the salience of the issue of stake (Clare 2007), the existence of an independent media or whistleblowers (Potter and Baum 2014), and by the leader's access to new information about the situation (Levendusky and Horowitz 2012). Existing research, however, has not examined whether the degree to which a leader follows through with her threat matters for audience cost theory.

Policy Substitution

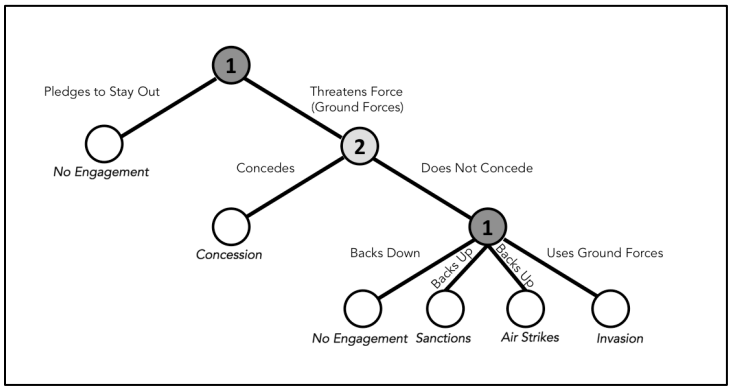
Before implementing a foreign policy, leaders generally consider a range of possible options, rather than developing just one potential course of action. Research on policy substitution suggests that several different foreign policy options can be used to address the same geopolitical issue (Most and Starr 1984; Palmer and Bhandari 2000; Clark and Reed 2005). For instance, when deciding how to counter a rival, a leader might choose between going to war (Gilpin 1981), strengthening alliances (Walt 1990), levying sanctions, or conducting internal balancing by increasing arms production. The ultimate decision on which policy to adopt is shaped in large part by the leader's assessment of the costs and benefits for each possible strategy, taking into account a host of domestic and international factors. Existing audience cost research has touched upon policy substitution, but does so by varying the initial threat or demonstration of force rather than analyzing different degrees of "backing up" once an initial threat has been made (Tomz 2007).

While the literature on policy substitution focuses on the assessment of a range of possible strategies *before* a foreign policy decision is made, a leader could arguably fall back on an alternate policy option *after* threatening another strategy. To be sure, critics might suggest that backing up is unlikely because leaders avoid making unambiguous threats that commit them to specific actions (G. H. Snyder and Diesing 1978; J. Snyder and Borghard 2011). Even ambiguous threats about "red lines" and "grave consequences", however, can strongly imply certain types of actions, and real world leaders have backed up from these non-explicit threats. Indeed, leaders may actually have incentives to bluff before backing down. An initial threat of hawkish military action might be intended to obtain concessions from a rival or respond to domestic political pressure. Leaders may subsequently back up to a different strategy if they receive new information about the situation (Levendusky and Horowitz 2012) or if an alternate policy provides a less risky or costly means of pursuing a strategic objective. In short, a leader may be willing to accept some audience costs if those costs are lower than the costs imposed for backing down entirely, or lower than the political costs that might be imposed if following through entirely with a threat generates significant casualties or international condemnation.

This crisis bargaining model with policy substitution is depicted in the stylized game in Figure 2, which demonstrates the increased strategy set available to a state leader who is willing to back up. The players are the same as in the classic crisis bargaining model, as are the strategies each player can adopt in their first move. However, if State 2 does not concede to State 1's threat, the policy substitution variant of the model expands the strategy set for State 1's second move. Rather than choosing only between "Backs Down" and "Uses Ground Forces", State 1 now has two additional policy options, which I label "Backs Up". In one "Backs Up"

strategy, State 1 launches airstrikes against State 2. In the other, State 1 implements sanctions. In an actual crisis, the strategy set might be far larger with several means of backing up.

FIGURE 2: Crisis Bargaining Model with Policy Substitution



The model I propose is informed not only by academic theories, but also by the recent practices of state leaders. As described earlier, President Obama publicly issued a threat to the Syrian government in August 2012, informing the Bashar al-Assad regime that the use of chemical weapons would be "a red line" that would change Washington's calculus on the use of military force. Although the statement avoided making an explicit threat of force, it strongly implied a threat of military action. Indeed, even members of Obama's staff believed that a Syrian violation of the red line would lead the United States to use force against the Assad regime (Chollet 2016, 8–26). One year later, Syrian government forces allegedly carried out chemical attacks on several opposition-controlled suburbs around Damascus, killing between 281 and 1,729 people (Mazzetti and Landler 2013). In response, the U.S. government began preparing for military strikes. Military action, however, never materialized as President Obama decided to "explore another option", even though the attacks clearly crossed his 2012 "red line." Instead of following through on his initial threat, Obama arguably backed up by agreeing to a multinational agreement under which the Syrian regime turned over its chemical weapons stockpiles for destruction (Chollet 2016, 8–26).

By considering "backed up" options – such as sanctions and airstrikes – rather than simply the conditions when a leader either follows through on a threat of force or backs down completely, this project attempts to capture real world dynamics more accurately than previous audience cost research. Introducing policy substitution to the study of audience costs allows us to gain a more nuanced understanding of whether the degree of inconsistency between words and deeds shapes how the public punishes their leaders.

Inconsistency as a Continuum

Although audience costs stem from the inconsistency between a leader's threats and actions, most literature on the topic has adopted an overly simplistic view of inconsistency, treating it as a binary variable. Literature on cognitive processes from the field of psychology, however, suggests that inconsistency can be measured along a continuum.[3] Individuals experience cognitive inconsistency and dissonance when a set of cognitive elements – pieces of information, beliefs, and opinions – diverge. Divergence may stem from a variety of factors including a diminishing correspondence between an individual's fundamental beliefs and a policy option (Converse 1964), or a decreasing level of qualitative similarity between one preference and an alternative option (Brehm and Cohen 1959). This divergence of cognitive elements occurs along a continuous spectrum in which the increasing discrepancy between one element and another generates increasing amounts of dissonance (Brehm and Cohen 1962, 11–17). When applied to divergence between a leader's words and deeds, certain types of actions may appear more consistent with an initial threat than others.

[3] Much of this literature concentrates on dissonance between actions and an individual's personal preferences, but a similar logic of dissonance can be applied to divergence between alternate policy options. For an example, see Brehm and Cohen 1959.

Since individuals generally favor consistency over inconsistency, they should prefer alternative goods or policies that minimize divergence from the goods or policies they initially hoped to obtain. Indeed, experimental studies have unsurprisingly shown that people prefer alternatives that are qualitatively similar to their most favored option (Brehm and Cohen 1959). When these logics of consistency and similarity are applied to the realm of audience costs, the public should prefer an action that is qualitatively similar in terms of perceived risks, costs, and military intensity to an initially promised action over an action that is qualitatively dissimilar.

How Backing Up Shapes Audience Costs and Crisis Signaling

If the preference for consistency over inconsistency is applied to audience cost theory, leaders should be able to reduce the punishment for shying away from threats. By opting to back up to actions that are qualitatively similar in risk, cost, or military intensity to the initially threatened policy, instead of backing down altogether, the leader potentially reduces the perceived inconsistency between his initial words and deeds. Since consistency is preferred to inconsistency, audience costs should be lower when the implemented action is more similar and consistent with initially threatened action. For instance, after a leader pledges an invasion, implementing another military action such as airstrikes might be viewed as being more consistent with the initial threat than taking only diplomatic measures or taking no action at all. This leads to two testable hypotheses on inconsistence and audience costs:

> H_1: *Domestic audiences will more favorably judge (i.e. impose lower audience costs on) a leader that "backs up" and implements a less hawkish intermediate policy than on a leader who backs down entirely.*

H_2: *As the qualitative difference (in risk, cost, or military intensity) between the initial threat and the implemented policy decreases, so too will the perceived inconsistency of the President's actions.*

Although the subject of an academic debate, many scholars suggest that the inconsistency associated with bluffing and reneging on threats generates reputational harm for leaders and states (Copeland 1997; Huth 1997). States with a track record for backtracking on threats are less likely to be trusted by allies and more prone to be challenged by rivals. Other studies, however, suggest that reputation plays little role in influencing the decisions of leaders and the outcomes of international crises (Mercer 1996). If the public believes that reneging on threats harms a state's reputation, the degree to which a leader's actions diverge from her words might shape the public's perception of the amount of reputational harm that is incurred by backing up. If the public believes adopting an intermediate action has a less adverse impact on a nation's ability to credibly issue future threats and promises than doing nothing at all, domestic audiences may judge a leader who backs up more favorably than one who backs down.

H_3: *The public will view a leader who backs up as having a less adverse impact on the state's reputation than a leader who backs down.*

At the same time, the public may view a leader who does something instead of nothing as more competent than a leader who reneges, even if the action taken differs from the initially issued threat. Leaders who go back on a threat can be seen as incompetent and not trustworthy (Guisinger and Smith 2002), and can have difficulty promoting their broader policy agenda after they poorly manage an interstate dispute (Gelpi and Grieco 2015). By backing up instead of reneging entirely, a leader is able to

claim to have taken steps toward a policy objective rather than having to explain a lack of action. In some cases, a leader may even be able to claim the backed up action was superior to the initially threatened option. Indeed, in describing his decision to back up from airstrikes, President Obama said he was "very proud of this moment…and I believe ultimately it was the right decision to make (Goldberg 2016)."

H_4: *The public will view a leader who backs up as more competent than a leader who backs down.*

METHODOLOGY

I test these hypotheses using an original survey experiment of 1,006 adult respondents in the United States. Because government leaders seek to avoid audience costs, they strategically select out of issuing empty threats, making audience costs difficult to study using observational data (Tomz 2007). In recent years, experimental methods have become increasingly common in international relations research and offer researchers a useful tool for studying subjects like audience costs where observational data is hard to collect (Hyde 2015).

Although survey experiments allow researchers to precisely identify causal effects by randomizing and controlling the application of treatments, they do not overcome all research challenges. Most significantly, there is much debate surrounding the external validity of experimental approaches (Hyde 2015). In other words, do the findings gathered under tightly controlled experimental conditions mean anything in the real world? Some critics of audience cost theory argue that external validity is limited as experimental studies overstate the public backlash that real world leaders would face for backing down (Clare 2007; Kurizaki and Whang 2015). Other critiques suggest that the artificially constructed scenarios in audience cost

experiments are unrealistic as they often feature explicit "committing threats" that actual leaders rarely make (J. Snyder and Borghard 2011).

While survey experiments are unable to perfectly mirror reality, they allow researchers to measure the causal effect of a certain variable – in this case, varying levels of policy inconsistency (Gerber and Green 2012, 1–17). International relations experiments are grounded in the plausible assumption that experimental subjects will apply conceptual and cognitive processes in a manner similar to non-experimental subjects in the real world (Schelling 1961, 55; Levendusky and Horowitz 2012, 328).

Experimental Design

My survey experiment draws heavily from the setup introduced in Michael Tomz's (2007) audience cost study, and exposes respondents to a scenario involving the U.S. President's decision to employ military force. All respondents are told that, "A country sent its military to take over a neighboring country in order to get more power and resources. The leader of the state being invaded asked the United States for assistance." Respondents are then randomly assigned into one of five experimental conditions that varies whether the President initially threatened to use force and whether the President followed through on his initial threat (experimental design in Figure 3, below). In one group the President pledges to stay out of the conflict and takes no action. In the other conditions, the President pledges to respond to the invasion using ground forces, but I vary whether he follows through with this threat, backs up by substituting an intermediate policy of sanctions or airstrikes, or backs down entirely and takes no action.[4]

In the control condition respondents were told, "The U.S. President said the United States would stay out of the conflict. The attacking country continued to invade. The U.S. President did

[4] Appendix A includes full text of the treatments and the survey instrument.

not send troops. In the end, the attacking country took over 20-percent of its neighbor's territory. No Americans were killed." In all four treatment conditions respondents were told, "the U.S. President said that if the attack continued, the U.S. military would use ground forces to push out the invaders." I then vary the degree to which the President subsequently follows-through with his initial threat. In the first treatment group, "the U.S. President deployed ground troops against the attacking country." In the second treatment group, "the U.S. President launched air strikes against the attacking country instead of sending ground troops." In the third treatment group, "the U.S. President launched economic sanctions against the attacking country instead of sending ground troops." And, in the fourth treatment group, "the U.S. President did not send troops."

TABLE 1: Experimental Design

	Follows Through	*Backs-Up 1*	*Backs-Up 2*	*Backs Down*
No Threat	Stays Out *n*=208	n/a	n/a	n/a
Threatens Ground Forces	Sends Ground Force *n*=198	Airstrikes *n*=208	Economic Sanctions *n*=195	No Military Action *n*=197

Note: The vertical access represents the President's initial decision on whether to threaten invasion. The horizontal axis represents the subsequent behavior (i.e. whether to follow through, back up, or back down). *n* represents the number of respondents randomly assigned into each experimental condition.

To avoid priming respondents, I avoid using potentially loaded phrases such as the President "followed through with" (or "reneged on", "backed down from", etc.) his threat in the vignettes presented in the survey instrument. The treatments also include no specific countries other than the United States to avoid influencing respondents' opinions. To control for policy outcomes, the attacking country "continued to invade" and "took over 20-percent of its neighbor's territory", regardless of the President's statements or actions. To control for opinions based on U.S. casualties, "No Americans were killed" in any of the experimental conditions.

Respondents are then asked for their opinion of the U.S. President's handling of the situation, which they rate on a five-point Likert scale ranging from strongly disapprove to

strongly approve. To better identify respondents' decision-making logic and to assess which underlying mechanism associated with audience cost logic underpins the experimental findings, the survey instrument also asks respondents several questions about their perceptions about consistency, reputation, and Presidential competence in the context of the hypothetical scenario. I also collect data on a variety of covariates including demographic data and information about respondent views on the use of force. These data allow me to compare the experimental sample to nationally representative samples, and to assess whether the treatment has heterogeneous effects based on demographic conditions.

To ensure subjects received and internalized the intended treatment, the survey instrument asks respondents to recall the action taken by the President. Of those assigned to the "stays out" treatment, 98-percent correctly recalled that the President "stayed out of the conflict". Among respondents in the "sends ground force" condition, 94-percent correctly identified that the President "deployed ground troops". In the "airstrikes" condition, 88-percent accurately remembered the President had "launched airstrikes." Among respondents assigned to the "sanctions" treatment, 85-percent accurately recalled that the President "enacted sanctions." Finally, 89-percent of respondents in the "backs down" condition recalled that in the end, the President "stayed out of the conflict."

Sampling and Implementation Strategy

The survey experiment was fielded on 1,006 adults across the United States in December 2016. Respondents were recruited using Amazon Mechanical Turk (MTurk) and randomly assigned into treatment groups using Qualtrics, the online platform used to implement the survey. Social scientists have increasingly turned to MTurk as a means of gathering data more quickly and inexpensively than traditional survey methods (Berinsky, Huber, and Lenz 2012;

Goodman, Cryder, and Cheema 2013; Buhrmester, Kwang, and Gosling 2011). Although crowd-sourced labor markets like MTurk allow researchers to cheaply and easily recruit research subjects, MTurk samples lack the representativeness of national probability samples (Berinsky, Huber, and Lenz 2012). The MTurk sample used in this experiment underrepresents blacks, Hispanics, and households with higher incomes, and overrepresents younger people and individuals with at least a high school degree. The gender balance and balance of military veterans and non-veterans in the experiment's sample are similar to those in a national sample.[5]

Although MTurk samples are less representative than national probability samples, studies have used subjects recruited using MTurk to replicate results from well-known social science experiments on risk perception, and obtained results similar to those obtained from nationally-representative samples (Berinsky, Huber, and Lenz 2012). These findings suggest that the cognitive processes of MTurk respondents are similar to those of the broader American population. Because of this, MTurk samples have been used in recent audience cost experiments (Chaudoin 2014).

EXPERIMENTAL FINDINGS

Before assessing the effects of treatment, I first verify that the assignment of subjects into the five experimental conditions was random. Using ordinary least squares models, I predict assignment into the treatment groups as a function of covariates including age, gender, education levels, income, political ideology, political involvement, and veteran status. The models suggest that respondents were indeed randomly assigned to experimental conditions. Thus, any

[5] Appendix B compares demographic characteristics of the experimental sample with a recent national sample.

differences in outcomes of interest between the experimental groups can be attributed to the treatment they receive in the survey instrument.[6]

Average Treatment Effects: Audience Costs

To assess the effects of policy substitution on audience costs, I examine respondents' approval ratings of the President's handling of the hypothetical scenario. Specifically, I measure the percentage of respondents in each experimental condition who either strongly disapprove or disapprove of the President's handling and calculate the mean approval rating for the President's actions on a five-point scale. Like earlier audience cost research, I interpret decreased approval levels as the audience cost imposed on a leader. Because the experimental design holds constant policy outcomes (i.e. no Americans die and the invader seizes 20-percent of the territory in all treatment conditions), differences in approval levels can be attributed to differences in the President's handling of the scenario. The findings show strong support for hypothesis H_1: backing up results in significantly lower audience costs than backing down. In line with H_2, audience costs increase as the implemented action becomes less qualitatively similar to the initially threatened ground invasion.

As Table 2 shows, 22.1-percent of respondents in the "stays out" condition (in which the President pledges to stay out and takes no military action) either disapprove or strongly disapprove of the President's actions. This represents the highest mean approval rating of any of the treatment conditions (3.385 on a 5-point scale). When the President follows through with his threat to deploy ground forces, the percentage of respondents who disapprove of the President's handling of the situation increases to 24.2-percent (mean approval: 3.359). The slight increase in

[6] Appendix C includes regression results for these balance tests.

disapproval may be explained, in part, by what Kertzer and Brutger (2016) describe as a belligerence cost, the punishment imposed when a leader threatens to use force.

Approval ratings take a substantial and statistically significant hit when the President deviates from his initial threat. The largest decrease in approval occurs when the President issues a threat to deploy ground forces, but takes no military action at all. Within this "backs down" treatment group, 58.9-percent of respondents disapprove of the President's actions (mean approval: 2.538). In other words, respondents in this group are approximately 2.4 times more likely to disapprove of the President than respondents in the group where the President follows through on his initial threat.

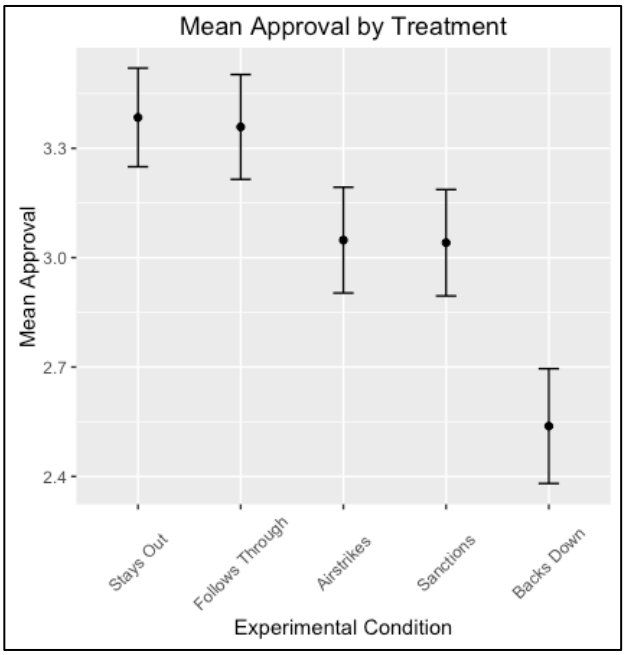
As Table 2 and Figure 3 depict, audience costs are not nearly as substantial when the President backs up to airstrikes or sanctions instead of backing down altogether. When the President backs up, his disapproval rating increases by roughly 12-percentage points over the follows through condition, compared to a 34.7-percentage point increase when he backs down. While a leader faces greater audience costs for backing up to sanctions than for backing up to airstrikes, the difference is far smaller than anticipated. These findings may suggest that the public views doing something as better than doing nothing, with little differentiation between backed-up options. The implementation of alternate policies – that lie between carrying out the full threat and doing nothing – appears to mitigate some of the audience costs associated with backing down.

TABLE 2: Effects of Policy Substitution on Approval

Experimental Condition	**Mean Approval Rating[1]**	**Change in Approval Rating**	**Respondents who Disapprove[2]**
Stays Out *Pledges to Stay Out, Takes No Military Action*	3.385 (.069)	--	22.1% (16.5-27.8)
Follows Through *Threatens Ground Forces, Sends Ground Forces*	3.359 (.073)	-0.026	24.2% (18.3-30.2)
Backs Up *Threatens Ground Forces, Launches Airstrikes*	3.048 (.074)	-0.337	36.1% (29.5-42.6)
Backs Up *Threatens Ground Forces, Implements Sanctions*	3.041 (.074)	-0.344	36.4% (29.7-43.2)
Backs Down *Threatens Ground Forces, Takes No Military Action*	2.538 (.080)	-0.847	58.9% (52.0-65.8)

1. Approval is measured on a five-point scale, where 1 is "strongly disapprove" and 5 is "strongly approve". Standard errors in parentheses. The effect of experimental conditions on approval ratings is significant at the p<.001 level.
2. Respondents who disapprove is a sum of the percentage of respondents who "Strongly Disapprove" and "Disapprove". 95-percent confidence interval in parentheses.

FIGURE 3: Approval Rates



To more systematically examine these findings, I employ a set of ordinary least squares models that allow me to test the statistical significance of findings and to assess whether there are heterogeneous treatment effects depending on respondent characteristics. All models regress

approval rating (measured on a five-point scale) on experimental condition (coded as a binary variable depending on whether a respondent was assigned to a given experimental group).

As Table 3 shows, Model 1 includes only the experimental conditions. The coefficients, which can be used to calculate mean levels of approval, reveal the effects of receiving each treatment. The intercept (2.538) represents the mean approval level when a respondent is assigned to the "backs down" condition. Approval increases by 0.50 points when the President chooses sanctions instead of backing down, by 0.51 points when the President opts for airstrikes and by 0.82 points when the President follows through on his threat. These findings support the conclusions of past audience cost research, but also show that intermediate "backed up" options provide a leader with the means to mitigate some audience costs associated with backing down altogether.

Model 2 adds a series of control variables such as age, education (coded on an eight-point scale ranging from less than high school to a postgraduate degree), political ideology (on a five-point scale from very liberal to very conservative), political involvement (whether the respondent participated in the 2016 Presidential election), attitudes toward intervention, and veteran status.[7] The coefficients on each of the treatment conditions remain comparable in size and in the same direction as those in the baseline model.

To assess whether the effects of treatment on approval ratings varied according to respondents' demographic characteristics, political ideology, and attitudes toward intervention, I tested several additional models that introduce interaction terms between the main variables of

[7] I introduce dummy variables for respondents who did not know or did not report their political ideology (n=12) and for respondents who were not eligible to vote in the 2016 Presidential election (n=16).

interest (i.e. experimental conditions) and various covariates.[8] Interactions with respondent gender, education level, political ideology, political involvement, and veteran status have no statistically significant effect on approval ratings. However, treatment effects do vary across different respondent attitudes toward intervention, as illustrated in Model 3 and Graph 4. To measure preferences toward intervention, the survey instrument asked respondents to rate (using a five-point scale from Strongly Disagree to Strongly Agree) their agreement with the statement "The United States needs to play an active role in solving conflicts around the world." Not surprisingly, respondents with more interventionist attitudes are, on average, likely to be less supportive of staying out (Coefficient: -0.297, p=.002) or backing down (Coefficient: -0.275, p<.001) than respondents with less interventionist perspectives. More interventionist respondents are also are more likely to support following through with an initial threat of ground forces (Coefficient: 0.694, p<.001) and airstrikes (Coefficient: 0.487, p<.001) than less interventionist subjects.

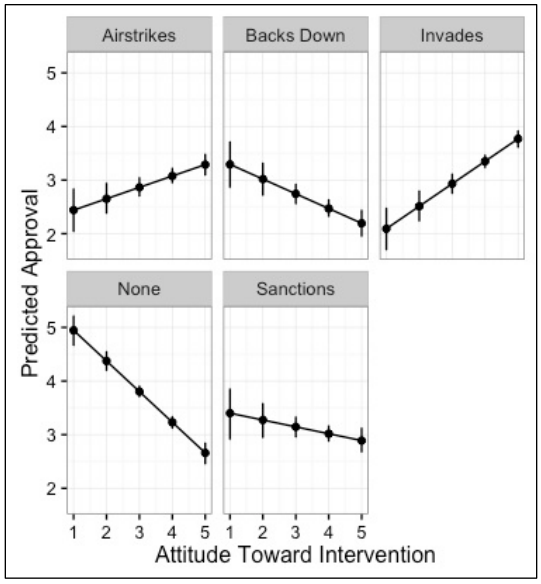
The fact that respondents reported their attitude toward intervention post-treatment limits the conclusions that can be drawn about the relationship between interventionist attitudes and approval of the President's actions. It is possible that exposure to an experimental condition influenced subjects' responses to a subsequent question about their support for intervention. As a result, findings about the heterogeneous effects of treatment across respondents with varying attitudes toward intervention is necessarily suggestive absent further testing.

[8] Appendix D includes regression tables for analysis of heterogeneous treatment effects.

Table 3: How Policy Substitution Shapes Approval

	Dependent variable:		
	Approval (OLS Coefficients)		
	(1)	(2)	(3)
Backs Down	2.538***	3.389***	3.292***
Threatens Ground Forces, No Military Action	(0.075)	(0.250)	(0.181)
Stays Out	0.847***	0.850***	1.654***
Pledges to Stay Out, No Military Action	(0.105)	(0.104)	(0.256)
Follows Through	0.821***	0.858***	−1.200***
Threatens Ground Forces, Sends Ground Forces	(0.106)	(0.106)	(0.262)
Backs Up: Airstrikes	0.510***	0.522***	−0.853***
	(0.105)	(0.104)	(0.253)
Backs Up: Sanctions	0.503***	0.523***	0.109
	(0.106)	(0.106)	(0.267)
Gender (Male)		−0.155**	
		(0.067)	
Age		−0.002	
		(0.003)	
Education		−0.073***	
		(0.022)	
Political Ideology (Conservative)		0.004	
		(0.031)	
Ideology (Don't Know)		−0.830***	
		(0.317)	
Political Involvement (Voted)		−0.023	
		(0.092)	
Political Involvement (Ineligible)		0.162	
		(0.275)	
Military Veteran		−0.007	
		(0.122)	
Interventionist Attitude		−0.050*	−0.275***
		(0.029)	(0.061)
Stays Out * Interventionist Attitude			−0.297***
			(0.086)
Follows Through * Interventionist Attitude			0.694***
			(0.084)
Airstrikes * Interventionist Attitude			0.487***
			(0.083)
Sanctions * Interventionist Attitude			0.147*
			(0.089)
Observations	1,006	1,006	1,006
R^2	0.077	0.102	0.216
Adjusted R^2	0.074	0.090	0.209
Residual Std. Error	1.053 (df = 1001)	1.043 (df = 992)	0.972 (df = 996)
F Statistic	20.993*** (df = 4; 1001)	8.655*** (df = 13; 992)	30.554*** (df = 9; 996)
Note:	Robust standard errors in parentheses. *p<0.1; **p<0.05; ***p<0.01		

FIGURE 4: Effect of Treatment on Approval Rating by Attitudes Toward Intervention



1. Approval is measured on a five-point scale, where 1 is "strongly disapprove" and 5 is "strongly approve." Attitude toward Intervention is measured on a five point scale where 1 is "strongly disagree" and 5 is "strongly agree."

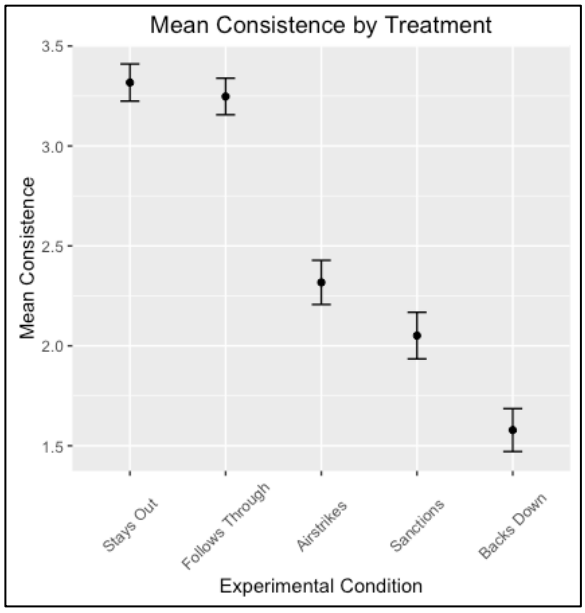
Average Treatment Effects: Consistency, Reputation, and Competence

To take a first cut at examining whether respondents judge a leader's consistency along a continuum, the survey instrument asks respondents "how consistent was the President's handling of the situation with his initial statement?" In line with our expectations (Hypothesis H_2), respondents perceive policies as more consistent when they diverge little from the initially threatened action. As Table 4 and Figure 5 illustrate, a leader who follows through on his threats or promises – either by honoring a pledge to stay out or a threat to send ground forces – receives relatively high consistency ratings. Unsurprisingly, consistency ratings decrease as a leader backs up or backs down from a threat. The consistency ratings reveal that the public perceives backing up to another military option (airstrikes) as more consistent with the initial threat than backing up to a non-military policy alternative (sanctions). This suggests that perceptions of consistency decrease as the qualitative divergence increases between a leader's words and deeds. When examined in conjunction with the audience cost measures above, more consistent behavior is generally correlated with higher levels of approval (i.e. lower audience costs).

Interestingly, while a leader who implements sanctions is seen as considerably less consistent than one who launches airstrikes, the President suffers roughly equal audience costs in the two backed up experimental condition (i.e. airstrikes and sanctions). These findings suggest that the public considers other factors, such as each policy's costs or operational risk, when deciding how to rate the President's actions. For instance, even if a respondent views airstrikes as being more consistent with an initial threat of ground invasion than sanctions, her level of disapproval over a choice to implement sanctions might be tempered if she believes the risks and costs associated with enacting sanctions are lower than those of conducting airstrikes. The microfoundations discussed below shed some light on this, but additional manipulations in the experimental design are required to assess how factors beyond inconsistency influence judgments about a leader's handling of crisis bargaining. These manipulations are left to future research.

TABLE 4 AND FIGURE 5: Consistency Ratings

Experimental Condition	**Mean Consistency Rating** [1]
Stays Out *Pledges to Stay Out, Takes No Military Action*	3.317 (0.047)
Follows Through *Threatens Ground Forces, Sends Ground Forces*	3.247 (0.047)
Backs Up *Threatens Ground Forces, Launches Airstrikes*	2.317 (0.056)
Backs Up *Threatens Ground Forces, Implements Sanctions*	2.051 (0.059)
Backs Down *Threatens Ground Forces, Takes No Military Action*	1.579 (0.055)



1. Consistency is measured on a four-point scale, where 1 is "Very Inconsistent" and 4 is "Very Consistent". Standard errors in parentheses. The effect of treatment on consistency ratings are significant to the p<.001 level.

According to the logic of audience costs, the public punishes leaders that renege on threats because inconsistency damages a state's reputation in future bargaining situations and because bluffing leaders are seen as weak or incompetent (Fearon 1994; Guisinger and Smith 2002). Since the findings described earlier suggest that the public ascribes different levels of inconsistency to different "backed-up" or "backed-down" policies, I use data about respondent perceptions of reputational consequences and leader competence to test hypotheses H_3 and H_4, assessing whether the degree of inconsistency between a leader's words and deeds results in differing views on reputational consequences and Presidential competence.

To solicit feedback on reputational effects, the survey instrument asked respondents to rate "How much did the President's handling of the situation hurt or improve the reputation of the United States in the world?" on a five-point scale ranging from "Significantly Hurt" to "Significantly Improve." Because reputation can mean different things to different people, I draw from Levy et al. (2015, 997) and ask a question about specific reputational consequences: "Based solely on the U.S. President's handling of the situation, how likely is it that other countries will believe threats and promises made by the U.S. President in the future?" Respondents answer on a four-point scale ranging from "Very Unlikely" to "Very Likely."

In line with hypothesis H_3, the public views leaders who back up as having a less adverse impact on the state's reputation than a leader who backs down. As Table 5 and Figures 6 and 7 indicate, the President's reputation and the credibility of future threats and promises is highest when the President follows through on his threat and lowest when the President backs down entirely. Launching airstrikes results in, on average, higher mean reputation and future threat ratings than enacting sanctions. These findings suggest that the public believes foreign allies and adversaries react differently to varying degrees of inconsistency between a President's words and

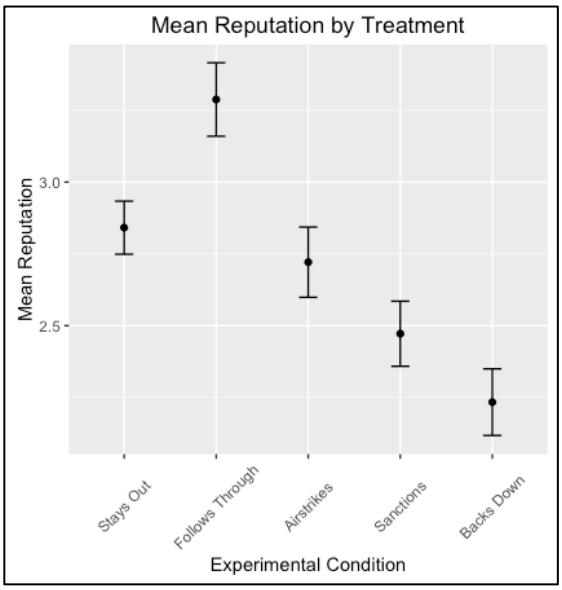
deeds in the foreign policy arena. Interestingly, the mean reputation and future threats and promises ratings for the "Stays Out" condition are approximately 0.4-points lower than those for the "Follows Through" condition, suggesting that the public believes that not coming to the aid of an ally in need weakens Washington's foreign policy reputation.

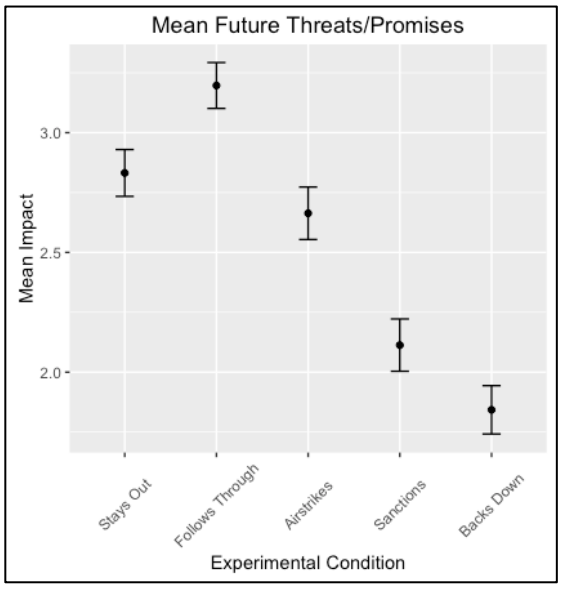
TABLE 5: Reputational Ratings

Experimental Condition	**Mean Reputation Rating[1]**	**Mean Future Threats/Promises Rating[2]**
Stays Out *Pledges to Stay Out, Takes No Military Action*	2.841 (.047)	2.831 (.050)
Follows Through *Threatens Ground Forces, Sends Ground Forces*	3.288 (.065)	3.197 (.049)
Backs Up *Threatens Ground Forces, Launches Airstrikes*	2.721 (.062)	2.663 (.056)
Backs Up *Threatens Ground Forces, Implements Sanctions*	2.472 (.058)	2.113 (.056)
Backs Down *Threatens Ground Forces, Takes No Military Action*	2.234 (.059)	1.843 (.051)

1. *Reputation* is measured on a five-point scale, where 1 indicates the President's handling of the situation will "Significantly Hurt" the reputation of the United States in the world, and 5 indicates his handling will "Significantly Improve" the reputation of the United States. Standard errors in parentheses. The effect of treatment on reputation rating are significant to the p<.001 level for all treatment conditions, with the exception of sanctions, where p=0.004.
2. *Future Threats* is measured on a four-point scale, where 1 indicates the President's handling of the situation will make it "Very Unlikely" that other countries will believe threats and promises made by the U.S President in the future, and 4 indicates that his handling make it "Very Likely" that other countries will believe the President's threats and promises in the future. Robust standard errors in parentheses. The effect of treatment on future threats rating are significant to the p<.001 level for all treatment conditions.

FIGURE 6 and 7: Reputational Ratings





Similar to the findings on approval, the experimental treatments have heterogeneous effects on reputation depending on the degree to which a respondent believes the United States should play an active role in solving conflicts around the world. Respondents with more interventionist attitudes are more likely to think that following through on a threat to deploy ground forces strengthens a state's reputation when compared to less interventionist respondents (Coefficient=0.365, p<.001). More interventionist subjects also view airstrikes as boosting the reputation of a state more than less interventionist subjects, although not as much as carrying out a threat of ground operations (Coefficient=0.229, p=.007). Not surprisingly, more interventionist respondents believe that pledging to stay out of the conflict harms a state's reputation (Coefficient=-0.167, p=.03) and the credibility of future threats and promises (Coefficient=-0.197, p=.004) more than less interventionist subjects.[9]

The data also reveal that the public views a leader who backs up as more competent than a leader who backs down, providing support for hypothesis H_4. Similar to the findings on reputation, the highest mean competence ratings are seen when the President follows through with his threat of ground forces and lowest when the President backs down entirely and takes no military action. The experimental manipulations also have statistically significant heterogeneous effects on competence ratings depending on respondent attitudes toward foreign intervention. Compared to their less interventionist counterparts, respondents with more interventionist leanings consider a President who pledges to stay out of the conflict as less competent (Coefficient=-0.248, p=.003). More interventionist respondents also offer higher competence ratings to leaders who follow through with some sort of military action (For ground forces:

[9] Appendix D includes regression tables for analysis of heterogeneous treatment effects. Findings based on interventionist attitude are solely suggestive for the reasons listed on page 21.

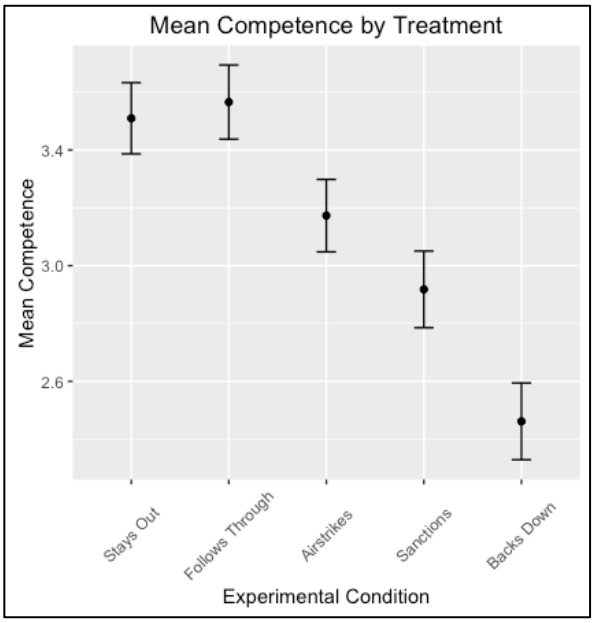
coefficient=0.516, p<.001. For airstrikes: coefficient=0.273, p=.002) than their less interventionist counterparts.[10]

TABLE 6: Presidential Competence Ratings

Experimental Condition	**Mean Competence Rating**[1]
Stays Out *Pledges to Stay Out, Takes No Military Action*	3.510 (.063)
Follows Through *Threatens Ground Forces, Sends Ground Forces*	3.566 (.065)
Backs Up *Threatens Ground Forces, Launches Airstrikes*	3.173 (.064)
Backs Up *Threatens Ground Forces, Implements Sanctions*	2.918 (.068)
Backs Down *Threatens Ground Forces, Takes No Military Action*	2.462 (.068)

1. *Competence* is measured on a five-point scale, where 1 indicates the President is "Very Incompetent" and 5 indicates he is "Very Competent". Standard errors in parentheses. The effect of experimental conditions on competence are significant to the p<.001 level.

FIGURE 8: Presidential Competence Ratings



[10] Appendix D includes regression tables for analysis of heterogeneous treatment effects on competence. Findings based on interventionist attitude are solely suggestive for the reasons listed on page 21.

Microfoundations: Explaining Experimental Outcomes

I turn to microfoundations – factors that "explain outcomes at the aggregate level via dynamics at a lower level" – to more satisfyingly examine the experimental findings (Kertzer 2017, 83). Audience cost theory rests upon micro-level assumptions, particularly individuals' distaste for inconsistency, to explain state behavior. To unpack whether concerns about inconsistency underpin approval for the President's handling of the hypothetical crisis, I examine survey experiment responses at the individual level of analysis in a manner similar to earlier studies (Tomz 2007). The findings of this microfoundations assessment support the inconsistency mechanism: respondents appear to punish leaders for being inconsistent. Further, respondents view inconsistency along a continuum. They are more likely to cite inconsistency as their justification for decreased approval when the implemented actions are more qualitatively different from the President's initial threat of ground forces.

To identify the factors that drove respondent approval ratings, I draw from data collected through an open-ended question in the survey instrument that asked respondents to "write a sentence or two telling us why you approve/disapprove of the way the President handled the situation." I then manually coded the responses of the 600 subjects assigned to the airstrikes, sanctions, and backs down conditions into one of five categories to assess *why* respondents gave lower approval ratings to leaders who failed to carry out their initial threat.[11]

One group of respondents believed the President did the right thing. Some of these respondents said the President's decision helped to save American lives, helped protect an ally,

[11] Since I am interested in the reasons respondents disapprove or approve of the President's backed-up and backed-downed actions, I code only respondents assigned to experimental conditions where the President reneged from his initial threat.

or simply because "it seemed to get the job done." A second group of respondents disapproved of the actions because they were opposed to international involvement or any use of force. Respondents in this category justified their positions with statements such as, "I think we need to stick more to protecting our own. We don't need to interfere in everyone else's conflicts[,]" or, "I don't approve of war. We need to stop killing each other." The third group of respondents held their positions because of the President's inconsistence. Some subjects believed that reneging on a threat could harm America's reputation by "mak[ing] it look like we do not honor our agreements" and because "foreign enemies will not take threats by the United States seriously." Others suggested that the inconsistence highlighted the President's weakness or lack of honor. A fourth group of respondents cited Presidential incompetence or poor planning without mentioning inconsistency. A fifth group stated that they had insufficient information or lacked the expertise to make a valid judgment.

Many respondents provided only a single justification for their approval rating. In some cases, however, respondents cited inconsistency in addition to another reason. For instance, some respondents suggested that they were opposed to government interference in the affairs of other states and supported the President's decision to withhold ground forces, but that they "[could not] strongly approve because the President said he would commit ground troops and he didn't follow through." The fact that many subjects tempered their approval rating despite their desire to avoid armed conflict or international intervention indicates the public will still hold leaders accountable for inconsistency. In cases where respondents offered multiple explanations, I coded the response into the inconsistency category if inconsistence was mentioned. If inconsistence was not mentioned, I coded the response into a single category based on the justification that was featured more predominantly in the respondent's free-text answer. Although manual coding is

inherently subjective, a reliability test involving an additional coder yielded an 87.3-percent inter-coder reliability rate.[12]

The results provide support for the commonly theorized microfoundations of audience costs. Like earlier studies, respondents in the "backs down" experimental condition (in which the President issues a threat to deploy ground forces then takes no action) are, on average, the most likely to justify their disapproval by citing concerns about inconsistency between a President's threats and actions. More importantly, the data offer greater insight into how domestic publics perceive inconsistency, something absent from earlier experimental research on audience costs. By examining justifications for approval ratings of "backed up" intermediate policies, I find that variation in the degree of inconsistency underpins the varying levels of Presidential approval between treatment conditions. As predicted by Hypothesis H_2, respondents in the two "backs up" conditions are, on average, less likely to reference inconsistency than respondents in the "backs down" condition. Table 7 shows that 61.9-percent of respondents assigned to the "backs down" condition mention inconsistency, compared to 39.5-percent when the President backs up to sanctions and 28.4-percent when the President backs up to airstrikes.

[12] A second individual coded all 600 responses and achieved an 87.3 percent agreement rate with the initial coder. Appendix E includes the coding guide used by both coders.

TABLE 7: Microfoundations by Experimental Condition

Justification for Approval Rating	**Backs Up** *Threatens Invasion, Implements Air Strikes*	**Backs Up** *Threatens Invasion, Implements Sanctions*	**Backs Down** *Threatens Invasion, Takes No Action*
Right Thing to Do	40.4 (33.7-47.1) n=84	30.8 (24.3-37.2) n=60	10.7 (6.4-15.0) n=21
Opposed to International Interference	19.2 (13.9-24.6) n=40	12.8 (8.1-17.5) n=25	14.2 (9.4-19.1) n=28
Inconsistency	28.4 (22.2-34.5) n=59	39.5 (32.6-46.3) n=77	61.9 (55.1-68.7) n=122
Ineffective Plan/Incompetence	4.8 (1.9-7.7) n=10	7.2 (3.6-10.8) n=14	6.6 n=13 (3.1-10.1)
Insufficient Information to decide	7.2 (3.7-10.7) n=15	9.7 (5.6-13.9) n=19	6.6 (3.1-10.1) n=13

Note: Entries represent the percentage of respondents in each treatment condition coded into each justification category. Percentages within each treatment category may not add to 100 because of rounding. Range in parentheses is a 95-percent confidence interval.

These findings suggest that once an initial threat of invasion is made, the public views acts of policy substitution along a continuum of inconsistency. The proportion of subjects citing inconsistency as the justification for their approval ratings increases as the President's actions become more qualitatively different from the initially threatened use of ground forces. Similarly, the percentage of respondents who believed the President's actions were the "right thing to do" decrease as the implemented action becomes more qualitatively different from the deployment of ground forces. Indeed, some respondents explicitly described the divergence between the President's threat and his subsequent actions. Several subjects described airstrikes as following through with the President's initial threat, suggesting that some members of the public view little distinction between different means of force employment. Sanctions, however, were often described as being less consistent with the deployment of ground forces. One subject explained that sanctions were "too different from the action promised[,]" while another suggested sanctions

were "more of an economic slap on the wrist than anything…it wasn't really intervention as promised."

The microfoundations data also provide insight into why the public punishes leaders for being inconsistent, confirming the reputational mechanisms thought to underlie audience cost theory (Fearon 1994; Guisinger and Smith 2002). Many respondents feared that a leader's failure to follow through on promises harms the state's reputation among friends and foes, and paints the leader as untrustworthy or incompetent. Some respondents expressed concerns that not following through on promises would "make it look like we do not honor our agreements with other nations" and "deteriorate our relations with foreign allies." Other respondents believed that reneging on threats meant that "foreign enemies will not take threats by the United States seriously" and that "other nations will prey on that weakness." In addition to concerns about future state reputation, respondents also suggested that a "President who doesn't follow through on his promises is a weak leader" and is not "reliable" or "honorable."

The overall findings of the microfoundational analysis support existing audience cost research, but also reveal that the punishment the public imposes on leaders is shaped by the degree of deviation between the initially pledged action and the implemented action. The findings provide a more nuanced understanding of how the public perceives inconsistency in foreign policy and assesses its consequences on a state's reputation.

CONCLUSION AND IMPLICATIONS

This paper extends prior research on audience costs and crisis bargaining by introducing the concept of "backing up" as an alternative to "backing down." Rather than assuming a leader has a binary choice of either following through with an initial threat or backing down completely, I argue that real world leaders often practice policy substitution. They can adopt one

of several policies after issuing an initial threat to use force. By backing up to a policy that is less inconsistent than backing down altogether, a leader can mitigate the audience costs she faces for reneging on a threat. I test this theory by fielding an original survey experiment and find that a leader can significantly reduce the audience costs she faces by backing up to sanctions or airstrikes instead of backing down entirely from an initial threat to deploy ground forces.

Respondents view these intermediate backed-up actions as more consistent with the initial threat than completely reneging, and therefore punish their leaders less than they would punish a leader for backing down from a threat entirely. Most demographic factors – such as education, gender, veteran status, or political involvement and ideology – have no significant influence on shaping the audience costs respondents dole out on inconsistent leaders. Respondents with more interventionist attitudes, however, are more likely than subjects with less interventionist leanings to punish a leader for implementing actions that shy away from an initially threatened ground operation. While the experimental findings support the mechanisms underlying audience cost theory, they also raise questions about the common claim that audience costs lead to more credible crisis signaling. Since backing up allows a leader to mitigate the punishment for deviating from threats, leaders might be more willing to step back from a strong initial threat than existing audience cost theories predict.

Implications and Pathways for Future Research

There are significant implications for audience cost theory and crisis bargaining if leaders can mitigate the punishment they face for reneging on threats by backing up to intermediate – and less inconsistent – positions. If leaders realize that backing up reduces audience costs, they may be tempted to issue a strong threat to force an adversary to the bargaining table or to change its behavior, before resorting to a potentially less risky or costly means of pursuing a desired

outcome. The incentive to back-up from a threat may be particularly great if the alternate action can achieve the same policy objective as the one initially threated, or if the initially threatened action is particularly risky or unpopular on the world or domestic stage. The ability to reduce the punishment for reneging on threats challenges the notion that leaders susceptible to audience costs avoid bluffing and issue only credible threats. Instead of making only threats they intend to follow through on – as existing audience policy theories predict – states might be more prone to issue hawkish threats before backing up to an alternate policy.

As a result, a state's ability to back up may make it difficult for both allies and adversaries to distinguish between a bluff and a credible threat. Indeed, the experimental findings on the effect of backing up on the believability of a state's future threats and promises suggests that individuals are less likely to believe subsequent warnings and assurances when a leader's backed up actions are more qualitatively different than the initially threatened action. Given the inability to identify whether a threat is credible or simply cheap talk, states on the receiving end of threats may then be more willing to engage in escalatory behavior to test the sending state's resolve. These states may be more willing to engage in escalatory probes when a rival's previous backed up actions demonstrated significant qualitative divergence from threatened actions, believing that they will face fewer consequences for their actions. Further research, however, is needed to validate this claim.

Variation in the economic and military capacity across states suggests that some states will be more able to back up to intermediate policies than other states. For instance, a state with only ground forces is unable to back up to the use of airpower and a state with weak economic influence cannot credibly enact sanctions. States with the ability to leverage the full range of diplomatic, military, and economic power can more easily back up after an initial threat. The

ability of powerful states to more easily back up appears to counter the assertions of both audience cost theorists and world leaders like former U.S. Vice President Joe Biden who once claimed, "big nations can't bluff."[13] Indeed, the findings of this experiment suggest that distinguishing credible threats from bluffs may be particularly difficult when the issuing state is a powerful one.

The findings of this study suggest several pathways for future research. First, future projects might explore whether the audience cost mitigating effects of backing up are generalizable beyond the United States. Most experiments examining audience costs have been fielded on respondents in the United States, but studies have suggested that audience costs affect leaders in capitals other than Washington (Tomz 2007; Davies and Johns 2013), including in autocratic states (Weeks 2008). Replicating the experiment in other democratic and non-democratic contexts may shed light on how variation in inconsistency shapes audience costs and public perceptions of reputation and leader competence. Second, further research might examine the role of media and politicians in shaping perceptions of inconsistency. Recent research has found that explaining decisions to renege on threats can reduce audience costs (Levendusky and Horowitz 2012). Can these sources control public perceptions about the degree of inconsistency between a leaders words and deeds? Third, future projects employing either experimental or observational methods might examine how backing up affects the credibility of a state's future threats or promises among actors on the receiving end of that state's threats. Finally, observational approaches – particularly case studies informed by rich qualitative evidence – may

[13] Vice-President Biden made this comment at the 2013 Annual Conference of the American Israel Public Action Committee (AIPAC). See, http://www.aipac.org/act/attend-events/policy-conference/videos/2013/speeches/biden for the full text of the speech.

be able to shed light on the processes that leaders use to determine when and how to back up rather than back down.

References

Berinsky, Adam J., Gregory A. Huber, and Gabriel S. Lenz. 2012. "Evaluating Online Labor Markets for Experimental Research: Amazon.com's Mechanical Turk." *Political Analysis* 20 (3): 351–68. doi:10.1093/pan/mpr057.

Brehm, Jack W., and Arthur R. Cohen. 1959. "Re-Evaluation of Choice Alternatives as a Function of Their Number and Qualitative Similarity." *The Journal of Abnormal and Social Psychology* 58 (3): 373. doi:10.1037/h0040493.

———. 1962. *Explorations in Cognitive Dissonance*. New York: John Wiley & Sons.

Buhrmester, Michael, Tracy Kwang, and Samuel D. Gosling. 2011. "Amazon's Mechanical Turk A New Source of Inexpensive, Yet High-Quality, Data?" *Perspectives on Psychological Science* 6 (1): 3–5. doi:10.1177/1745691610393980.

Chaudoin, Stephen. 2014. "Promises or Policies? An Experimental Analysis of International Agreements and Audience Reactions." *International Organization* 68 (1): 235–56. doi:10.1017/S0020818313000386.

Chollet, Derek. 2016. *The Long Game: How Obama Defied Washington and Redefined America's Role in the World*. New York: PublicAffairs.

Clare, Joe. 2007. "Domestic Audiences and Strategic Interests." *The Journal of Politics* 69 (3): 732–45. doi:10.1111/j.1468-2508.2007.00571.x.

Clark, David H., and William Reed. 2005. "The Strategic Sources of Foreign Policy Substitution." *American Journal of Political Science* 49 (3): 609–24. doi:10.2307/3647735.

Converse, Phillip E. 1964. "The Nature of Belief Systems in Mass Publics." In *Ideology and Discontent*, edited by David Apter, 202–61. New York: Wiley.

Copeland, Dale C. 1997. "Do Reputations Matter?" *Security Studies* 7 (1): 33–71. doi:10.1080/09636419708429333.

Davies, Graeme A. M., and Robert Johns. 2013. "Audience Costs among the British Public: The Impact of Escalation, Crisis Type, and Prime Ministerial Rhetoric." *International Studies Quarterly* 57 (4): 725–37. doi:10.1111/isqu.12045.

Fearon, James D. 1994. "Domestic Political Audiences and the Escalation of International Disputes." *The American Political Science Review* 88 (3): 577–92.

Gelpi, Christopher, and Joseph M. Grieco. 2015. "Competency Costs in Foreign Affairs: Presidential Performance in International Conflicts and Domestic Legislative Success, 1953–2001." *American Journal of Political Science* 59 (2): 440–56. doi:10.1111/ajps.12169.

Gerber, Alan S, and Donald P Green. 2012. *Field Experiments: Design, Analysis, and Interpretation*. New York: W. W. Norton.

Gilpin, Robert. 1981. *War and Change in World Politics*. New York: Cambridge University Press.

Goldberg, Jeffrey. 2016. "The Obama Doctrine." *The Atlantic*, April. https://www.theatlantic.com/magazine/archive/2016/04/the-obama-doctrine/471525/.

Goodman, Joseph K., Cynthia E. Cryder, and Amar Cheema. 2013. "Data Collection in a Flat World: The Strengths and Weaknesses of Mechanical Turk Samples." *Journal of Behavioral Decision Making* 26 (3): 213–24. doi:10.1002/bdm.1753.

Guisinger, Alexandra, and Alastair Smith. 2002. "Honest Threats: The Interaction of Reputation and Political Institutions in International Crises." *The Journal of Conflict Resolution* 46 (2): 175–200.
Huth, Paul K. 1997. "Reputations and Deterrence: A Theoretical and Empirical Assessment." *Security Studies* 7 (1): 72–99. doi:10.1080/09636419708429334.
Hyde, Susan D. 2015. "Experiments in International Relations: Lab, Survey, and Field." *Annual Review of Political Science* 18 (1): 403–24. doi:10.1146/annurev-polisci-020614-094854.
Jervis, Robert. 1970. *The Logic of Images in International Relations*. Princeton, N.J.: Princeton University Press.
Kertzer, Joshua D. 2017. "Microfoundations in International Relations." *Conflict Management and Peace Science* 34 (1): 81–97. doi:10.1177/0738894216665488.
Kertzer, Joshua D., and Ryan Brutger. 2016. "Decomposing Audience Costs: Bringing the Audience Back into Audience Cost Theory." *American Journal of Political Science* 60 (1): 234–49.
Kurizaki, Shuhei, and Taehee Whang. 2015. "Detecting Audience Costs in International Disputes." *International Organization* 69 (4): 949–80.
Levendusky, Matthew S., and Michael C. Horowitz. 2012. "When Backing Down Is the Right Decision: Partisanship, New Information, and Audience Costs." *The Journal of Politics* 74 (2): 323–38.
Levy, Jack S., Michael K. McKoy, Paul Poast, and Geoffrey P.R. Wallace. 2015. "Backing Out or Backing In? Commitment and Consistency in Audience Costs Theory." *American Journal of Political Science* 59 (4): 988–1001.
Mazzetti, Ben Hubbard, Mark, and Mark Landler. 2013. "Blasts in the Night, a Smell, and a Flood of Syrian Victims." *The New York Times*, August 26. http://www.nytimes.com/2013/08/27/world/middleeast/blasts-in-the-night-a-smell-and-a-flood-of-syrian-victims.html.
Mercer, Jonathan. 1996. *Reputation And International Politics*. Ithaca, NY: Cornell University Press.
Most, Benjamin A., and Harvey Starr. 1984. "International Relations Theory, Foreign Policy Substitutability, and 'Nice' Laws." *World Politics* 36 (3): 383–406. doi:10.2307/2010380.
Palmer, Glenn, and Archana Bhandari. 2000. "The Investigation of Substitutability in Foreign Policy." *The Journal of Conflict Resolution* 44 (1): 3–10.
Potter, Philip B. K., and Matthew A. Baum. 2014. "Looking for Audience Costs in All the Wrong Places: Electoral Institutions, Media Access, and Democratic Constraint." *The Journal of Politics* 76 (1): 167–81. doi:10.1017/S0022381613001230.
Schelling, Thomas C. 1961. "Experimental Games and Bargaining Theory." *World Politics* 14 (1): 47–68.
Schultz, Kenneth A. 1998. "Domestic Opposition and Signaling in International Crises." *The American Political Science Review* 92 (4): 829–44.
Snyder, Glenn H., and Paul Diesing. 1978. *Conflict Among Nations: Bargaining, Decision Making, and System Structure in International Crises*. Princeton, N.J.: Princeton University Press.
Snyder, Jack, and Erica Borghard. 2011. "The Cost of Empty Threats: A Penny, Not a Pound." *The American Political Science Review* 105 (3): 437–56.

Tomz, Michael. 2007. “Domestic Audience Costs in International Relations: An Experimental Approach.” *International Organization* 61 (4): 821–40.
Walt, Stephen. 1990. *The Origins of Alliance*. Ithaca, N.Y.: Cornell University Press.
Weeks, Jessica L. 2008. “Autocratic Audience Costs: Regime Type and Signaling Resolve.” *International Organization* 62 (1): 35–64.