ESSAYS ON DOMESTIC POLITICS OF
INTERNATIONAL DISPUTES AND COOPERATION

RYAN DAVID BRUTGER

A DISSERTATION
PRESENTED TO THE FACULTY
OF PRINCETON UNIVERSITY
IN CANDIDACY FOR THE DEGREE
OF DOCTOR OF PHILOSOPHY

RECOMMENDED FOR ACCEPTANCE
BY THE DEPARTMENT OF
POLITICS
ADVISERS: KRISTOPHER RAMSAY AND HELEN MILNER

MAY 2016
© Copyright by Ryan David Brutger, 2016.

All rights reserved.
Abstract

The three essays that comprise this dissertation examine how domestic politics affect international disputes and cooperation. The first chapter examines dispute escalation at the WTO, presenting a theory of informational lobbying by firms for trade liberalization, not through political contributions, but instead through contributions to the litigation process. In this “litigation for sale” model, firms signal information about the strength and value of potential cases, and the government screens cases based on firms’ signals. The model suggests that firm participation increases a states’ ability to efficiently pursue the removal of trade barriers and helps explain the unusually high success rate for complainants in WTO disputes. The second chapter recognizes that conventional wisdom assumes that leaders who compromise lose the respect of their constituents and damage their reputations, which in turn undermines the prospects for peace and security. I present research that challenges this assumption and tests when and how leaders can negotiate compromises and avoid paying domestic costs. I argue that leaders avoid audience costs by exercising “proposal power” and initiating compromises. These contributions suggest that leaders who exercise proposal power have significant flexibility to negotiate compromise settlements. The third chapter of my dissertation engages a literature that argues one way governments can credibly signal their intentions in foreign policy crises is by creating domestic audience costs. This chapter argues that there are actually two logics of audience costs: audiences can punish leaders both for being inconsistent, and for threatening to use force in the first place. We employ an experiment that disentangles these two rationales, and turn to a series of dispositional characteristics from political psychology to bring the audience into audience cost theory. Our results suggest that traditional audience cost experiments may overestimate how much people care about inconsistency, and that the logic of audience costs (and the implications for crisis bargaining) varies considerably with the leader’s constituency.
Acknowledgements

The process of completing this dissertation was often a solo endeavor, but one that was realized in great part due to the support of countless people. First, I must thank my husband and son for providing me perspective, balance, and understanding as I progressed through the unrelenting journey of completing my PhD. Without their continued love and support, I don’t know where this journey would have taken me. I would also like to thank the rest of my family for their immense support, not only during my work on this dissertation, but for everything that came before and has led me here. Additionally, some of the most important parts of writing a dissertation occur when sharing thoughts over a pitcher of beer, venting about pitfalls with friends at conferences, and relaxing with a glass of wine, good friends, and bad TV. For all the friends who made these years a little brighter, thank you!

Of course, a dissertation would not be possible without the dedication and guidance of one’s advisors. For this, I am very grateful for all that my advisors have done for me. Thank you for giving me the opportunity to come to Princeton, for the thoughtful conversations and advice, and for always pushing me to be better. I also thank the multitude of other faculty members who are not officially my advisors, but nonetheless gave countless hours of their time to support my journey.

I would also like to thank Joshua Kertzer, who is a coauthor on Chapter 3 of the dissertation. It has been a joy to work with you, and I look forward to many more projects in the years ahead.

Lastly, I would like to thank the many friends, faculty, and anonymous reviewers who provided feedback on earlier versions of this work. Portions of this dissertation benefited from presentations at Dartmouth College, Princeton University, the University of Arizona, Yale University, University of Pittsburgh, Columbia University, the London School of Economics, NYU-Abu Dhabi, the University of Minnesota, the University of Pennsylvania, and conferences of the American Political Science Association, the Midwest Political Science As-
society, the Southern Political Science Association, the International Political Economy Society, and the Peace Science Society (International). I also am thankful for financial support from the Neihaus Center for Globalization and Governance, the Center for International Security Studies, the Fund for Experimental Social Sciences, and the Mamdouha S. Bobst Center for Peace and Justice, which made this research possible.
Contents

Abstract . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . iii
Acknowledgements . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . iv
List of Tables . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . x
List of Figures . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . xi

1 Screening for Success: The Effect of Firm Signaling on WTO Case Selection 1
   1.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 1
   1.2 Framing Dispute Settlement Participation . . . . . . . . . . . . . . . . . . 4
   1.3 The Argument . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 7
   1.4 The Model . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 11
   1.5 Evidence of Firm Signaling and Contributions . . . . . . . . . . . . . . . . 20
   1.6 Conclusion . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 30

2 When to Compromise? The Influence of Compromise, Proposal Power, and Partisanship on Domestic Audience Costs 33
   2.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 33
   2.2 Diplomacy, Compromise, and Audience Costs . . . . . . . . . . . . . . . . 36
   2.3 Reconsidering What Audiences Want . . . . . . . . . . . . . . . . . . . . . 40
   2.4 Experimental Methods . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 45
   2.5 Results . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 49
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.5.1 The Effect of Compromise</td>
<td>51</td>
</tr>
<tr>
<td>2.5.2 Sensitivity to Outcomes</td>
<td>52</td>
</tr>
<tr>
<td>2.5.3 Proposal Power</td>
<td>54</td>
</tr>
<tr>
<td>2.5.4 Signaling and Lock-In</td>
<td>55</td>
</tr>
<tr>
<td>2.6 Conclusion</td>
<td>57</td>
</tr>
<tr>
<td>3 Decomposing Audience Costs: Bringing the Audience Back Into Audience</td>
<td>60</td>
</tr>
<tr>
<td>3.1 Introduction</td>
<td>60</td>
</tr>
<tr>
<td>3.2 Audience Costs, Revisited</td>
<td>62</td>
</tr>
<tr>
<td>3.3 Audience Costs as a Double-Barreled Treatment Effect</td>
<td>64</td>
</tr>
<tr>
<td>3.4 Bringing the Audience Back In</td>
<td>70</td>
</tr>
<tr>
<td>3.5 Method</td>
<td>74</td>
</tr>
<tr>
<td>3.6 Results</td>
<td>77</td>
</tr>
<tr>
<td>3.6.1 Average Treatment Effects</td>
<td>77</td>
</tr>
<tr>
<td>3.6.2 Heterogenous Logics of Audience Costs</td>
<td>78</td>
</tr>
<tr>
<td>3.7 Implications</td>
<td>83</td>
</tr>
<tr>
<td>3.8 Conclusion</td>
<td>86</td>
</tr>
<tr>
<td>A Screening for Success Appendix</td>
<td>88</td>
</tr>
<tr>
<td>A.1 Summary of Key Variables:</td>
<td>88</td>
</tr>
<tr>
<td>A.2 Equilibrium Solution</td>
<td>89</td>
</tr>
<tr>
<td>A.2.1 Overview of Equilibrium Solution</td>
<td>89</td>
</tr>
<tr>
<td>A.2.2 Proof of Equilibrium Solution</td>
<td>90</td>
</tr>
<tr>
<td>A.3 Uniqueness of Equilibrium</td>
<td>98</td>
</tr>
<tr>
<td>A.3.1 Mixed Possibility</td>
<td>98</td>
</tr>
<tr>
<td>A.3.2 Alternative Semi-Pooling Possibilities</td>
<td>102</td>
</tr>
<tr>
<td>Section</td>
<td>Title</td>
</tr>
<tr>
<td>---------</td>
<td>----------------------------------------------------------------------</td>
</tr>
<tr>
<td>A.3.3</td>
<td>Alternative Separating Possibility</td>
</tr>
<tr>
<td>A.3.4</td>
<td>Pooling Possibility 1a</td>
</tr>
<tr>
<td>A.3.5</td>
<td>Pooling Possibility 1b</td>
</tr>
<tr>
<td>A.3.6</td>
<td>Pooling Possibility 2a</td>
</tr>
<tr>
<td>A.3.7</td>
<td>Pooling Possibility 2b</td>
</tr>
<tr>
<td>A.4</td>
<td>Multiple Firms with Incomplete Information</td>
</tr>
<tr>
<td>A.5</td>
<td>Robustness Checks</td>
</tr>
<tr>
<td>A.5.1</td>
<td>Sample Variation in Empirical Models</td>
</tr>
<tr>
<td>A.5.2</td>
<td>Industry Level Fixed Effects</td>
</tr>
<tr>
<td>B</td>
<td>When to Compromise Appendix</td>
</tr>
<tr>
<td>B.1</td>
<td>Sample Demographics</td>
</tr>
<tr>
<td>B.2</td>
<td>Comparing mTurk and Qualtrics Results</td>
</tr>
<tr>
<td>B.3</td>
<td>The Impact of Holding Casualties Constant</td>
</tr>
<tr>
<td>B.4</td>
<td>Average Approval Not Conditioned on Party of the President</td>
</tr>
<tr>
<td>B.5</td>
<td>Full Results for Experiments 1 and 2</td>
</tr>
<tr>
<td>B.6</td>
<td>Effects Among Independents</td>
</tr>
<tr>
<td>B.7</td>
<td>Mediation Analysis - Reputation and Support for Compromise</td>
</tr>
<tr>
<td>B.8</td>
<td>Experiment 2 - Treatments</td>
</tr>
<tr>
<td>C</td>
<td>Decomposing Audience Costs Appendix</td>
</tr>
<tr>
<td>C.1</td>
<td>Dispositional Measures</td>
</tr>
<tr>
<td>C.2</td>
<td>Sampling Methodology and Sample Characteristics</td>
</tr>
<tr>
<td>C.3</td>
<td>Alternative Experimental Designs</td>
</tr>
<tr>
<td>C.4</td>
<td>Simple Effects Versus Main Effects</td>
</tr>
<tr>
<td>C.5</td>
<td>The Impact of Specifying Casualty Levels on the Cost of Threatening Force</td>
</tr>
<tr>
<td>C.6</td>
<td>Regression Analysis</td>
</tr>
</tbody>
</table>
List of Tables

1.1 Random Effects Logistic Regression - Dominant Firm Capacity . . . . . . . . 25
1.2 Random Effects Logistic Regression - Average Firm Capacity . . . . . . . . 28

2.1 Experiment 1 - Differences in Approval Between Not Engage and Compromise 52
2.2 Experiment 2 - Differences in Approval Based on Outcomes . . . . . . . . . 54

3.1 Decomposing Audience Costs: Threatening Force Versus Inconsistency . . . 79

A.1 Random Effects Logistic Regression of WTO Dispute Complaints . . . . . . 120
A.2 Industry Fixed Effects Logistic Regression of WTO Dispute Complaints . . . 121

B.1 Sample Populations and National Population Comparison . . . . . . . . . 122
B.2 Comparison of Effects in mTurk Experiment 2 and Qualtrics Samples . . . 123
B.3 Experiment 1: Differences in Approval Score for Strategies and Party . . . 126
B.4 Experiment 2: Differences in Approval Score for Strategies and Party . . . 127
B.5 Experiment 1: Differences in Approval Score for Strategies and Party . . . 128

C.1 Survey Sample Characteristics . . . . . . . . . . . . . . . . . . . . . . . . . . 134
C.2 Regression Models . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 142
C.3 Decomposing Audience Costs (In Percentage Point Change in Disapproval) 146
List of Figures

1.1 Extensive Form Game ................................................. 14
1.2 Firm and Government Equilibrium Contributions ....................... 17
1.3 Change in Predicted Probability Plots .................................. 27

2.1 Crisis-Bargaining Model .................................................. 39
2.2 Average Approval Score for Each Strategy ................................. 50
2.3 Average Approval Score for Compromise .................................. 56

3.1 Classic Crisis-Bargaining Model .......................................... 63
3.2 Experimental Design and Revised Crisis-Bargaining Model ............... 65
3.3 The multiple logics of audience costs ....................................... 80

B.1 Average Approval Score for Each Strategy ................................. 125
B.2 Average Causal Mediation Analysis ....................................... 129

C.1 Comparison of the Probability Distributions for \[
\frac{E[Y_{22} - Y_{21}]}{E[Y_{22} - Y_{11}]}\] Estimated Using a Mean-Based Approach Versus OLS ..................... 143
Chapter 1

Screening for Success: The Effect of Firm Signaling on WTO Case Selection

1.1 Introduction

Given the consensus among economists that free trade is welfare enhancing, domestic interest groups are often blamed for the persistence of trade barriers. Yet even though “protection for sale” arguments have significant support (Gawande and Bandyopadhyay 2000; Goldberg and Maggi 2001; Grossman and Helpman 1994), domestic firms also play a prominent role in maintaining the liberal trading system and monitoring states’ international trade policies. In contrast to a significant body of work that examines when and why trade barriers arise, this paper studies how firms and governments monitor trade barriers and select which barriers to contest. While no agreement or institution has done more to liberalize the rules of the trading system than the General Agreement on Tariffs and Trade (GATT) and subsequently the World Trade Organization (WTO), states regularly impose barriers that are in conflict
with their WTO obligations. In the presence of a multitude of potentially noncompliant trade barriers, states must decide how best to allocate their resources to monitor and enforce trade agreements. Building from theories of informational lobbying (Chalmers 2013; Lohmann 1995; Potters and Winden 1991), this paper models the interaction between firms and their government and finds that a type of “litigation for sale” occurs. Unlike traditional models of lobbying, where interest groups make campaign contributions or offer election support, this paper identifies an alternative form of lobbying through litigation contributions – contributions to the fact-finding efforts, research costs, and litigation tasks – which play an important informational role by signaling the strength and value of potential trade disputes. In this manner, firms act as fire alarms (McCubbins and Schwartz 1984), allowing the government to screen cases and more efficiently monitor and enforce international trade agreements.

This paper examines the role and consequences of private firm participation in WTO dispute settlement, arguing that firms play an important role in monitoring WTO compliance and screening potential WTO complaints. Although the WTO restricts dispute initiation to national governments, I show that private firms play an active role in the dispute settlement process. The model presented here not only highlights the role of firms in maintaining the liberal trading system, but also contributes to burgeoning literatures in international relations on transnational versus interstate dispute settlement and the importance of formal versus informal rules of international organizations. I argue that the formal rules of the WTO Dispute Settlement Understanding allow its members to benefit from increased monitoring and enforcement provided by informal private firm participation, without governments taking on the additional risk associated with transnational dispute settlement, where private firms have direct legal access (Allee and Peinhardt 2010; Simmons 2014). Unlike their role in transnational dispute settlement mechanisms, where firms’ access to international arbitration can often put them at odds with their home government, I show that the WTO rules allow
governments to garner increased information and resources from firms, while preserving their role as legal gatekeepers. The result is that states are able to more efficiently screen and monitor potential WTO cases.

I show that private firms monitor WTO compliance and motivate states to seek enforcement of treaty obligations in two complementary ways. From a purely economic perspective, firms can contribute resources to support the litigation of WTO disputes, which reduces the costs of filing a complaint for the state. More importantly, firms are also positioned to signal information regarding the legal strength of potential cases, which allows the government to more accurately predict the probability of success. As the gatekeepers, governments screen cases based on potential strength and value, which helps explain the nearly 90 percent success rate of WTO complainants. I extend the analysis to examine firms’ incentives to monitor and seek enforcement of international legal obligations across industries. Moving beyond an analysis of just the firm and government, I employ a formal model to analyze when dominant firms within industries are likely to provide litigation contributions, overcoming free riding problems, and when firms are less likely to seek enforcement of trade obligations. The formalization of the argument also helps demonstrate why firms resist bluffing when cases are weak and highlights subtleties of the equilibrium not common in traditional signaling models. The implications of the model are consistent with qualitative and statistical evidence, suggesting that private firms use informal mechanisms to lobby for enforcement of states’ WTO obligations.

In the broader context of the international compliance and enforcement literature, this paper contributes to the debate over how and when non-state actors mobilize to encourage increased state compliance with international law. While many scholars agree that private actors play important roles in determining when states comply with and seek enforcement of
international law, how and when these actors change state behavior remains a contested issue (Bown 2009, Chap. 5; Chaudoin 2014a; Dai 2005; Davis 2012b; Johns and Rosendorff 2009). Focusing on the WTO, which is arguably the most influential international economic institution, I demonstrate the informational role of private firm mobilization on states’ selection of WTO complaints.

This paper proceeds in the following manner; the next section frames the debate over dispute settlement participation and discusses key determinants of case selection. I then develop a theory of firm participation, which centers on firms’ ability to alter the case selection process of states by signaling the potential legal strength and value of the case by contributing to the litigation process. The theory is then formalized in a model that highlights the important role of dominant firms within industries for the case selection process. The implications of the model are tested using firm-level data with a dataset of potential US trade disputes and are supported by interviews with international government officials and legal experts. Finally, the paper concludes with a discussion of some of the implications for theory and policy.

1.2 Framing Dispute Settlement Participation

Much of the existing discussion over WTO dispute settlement overlooks the role of domestic firms and instead examines other determinants of participation, which can be divided into research regarding which states choose to participate and which cases those states choose to bring to the WTO. It is generally agreed that a country will engage in a dispute when its expected benefits outweigh the expected costs (Bown 2005b). Understanding what factors states evaluate when considering the expected costs and benefits of a case is a critical step to delineating how private parties alter the cost-benefit calculation of states.

Significant research has focused on the costs of initiating a dispute. The direct costs associated with bringing a WTO dispute are often several million dollars. According to one
In addition to the direct costs of disputes, Horn and Mavroidis (2006) note that additional factors are often given causal weight in determining how states calculate the cost-benefit trade-off of WTO dispute settlement. One particular factor that has been examined in some detail is how power relations between potential disputants affect the probability a dispute is initiated. Research by Guzman and Simmons (2005) found that the “threat of coercive tactics by the powerful” state does not appear to be a major problem for WTO law. On the other hand, Bown (2005b) does find that states’ retaliatory capacity can play a significant role in determining whether a country files a WTO complaint.

Horn and Mavroidis (2006) also note that legal capacity and trade interest are important factors when states select potential complaints. The idea that a state’s trade interests motivate its decision to participate in a trade violation complaint is rather intuitive. Using trade interest as an explanatory variable seeks to account for the magnitude of injury a country experiences from a loss of market access derived from a non-competitive measure. Supporting this concept, Bown (2005a) concludes that the magnitude of exports affected by a particular case is positively correlated to the probability the case was initiated. This paper builds upon this and similar studies that support the understanding that states conduct a cost-benefit analysis when considering participation in WTO complaints (Horn, Mavroidis, and Nordstrom 1999; Bown 2005b).

The capacity of a state to engage in the dispute settlement process, including both the financial resources of a state and the state’s legal knowledge, also affects states’ participation.
in WTO disputes. Shaffer (2006) argues that two important limitations to states’ participation are a lack of legal expertise in WTO law and financial constraints to organizing effective representation in the WTO legal system. Empirical tests of the effect of litigation resources and previous WTO experience on dispute initiation have shown both to have positive effects (Bown 2005a; Davis and Bermeo 2009; Horn, Mavroidis, and Nordstrom 1999). Yet even among the WTO members least constrained by legal knowledge and resources, such as the US and the EU, the private sector often plays a role in relaxing these constraints, while also signaling the strength of the potential case. For example, in the disputes DS316/DS347 and DS317/DS353 between the United States (Boeing) and the European Community (Airbus) each firm hired legal representation for the dispute settlement process (World Trade Organization 2010). Boeing employed the law firm Wilmer Cutler Pickering Hale & Dorr to assist in the WTO subsidy case and Airbus employed Sidley Austin LLP as counsel for the case (Sidley Austin 2009; Wilmer Hale 2010). The estimated contributions from the private firms to the litigation budget “were running at $1,000,000 per month and could reach $20,000,000 for each company...” (Shaffer 2008, 184). The striking role of private firms in the Airbus-Boeing disputes illuminates the importance of firms in mitigating resource constraints, while also playing an important informational role.

I build upon existing theories of informational lobbying to assess the role of private firms in WTO dispute initiation, while also considering the role of litigation contributions to signal the credibility of information and mitigate resource constraints that limit WTO participation. Although the WTO limits formal case initiation to governments, I choose not to limit the analysis by assuming that states’ resources are the only relevant input and instead aggregate available resources, which include the information and financial resources of the state and private firms. Although the model examines the interaction between the firm and government and firms within an industry, the firm could also be a trade association,

---

5The EU is considered as a single entity because trade policy is centrally coordinated (Meunir 2005).
advocacy group, or other non-state actor with a vested interest and private knowledge of the case.

1.3 The Argument

Existing arguments regarding private firms’ influence on dispute settlement participation are generally limited to firms’ ability to define the trade agenda of states through traditional lobbying or government established mechanisms, such as Section 301 petitions in the United States (Bown and Hoekman 2005). Although these means of influence are significant, some firms will employ additional measures, specifically contributing to the litigation process in an effort to increase the likelihood a case is brought to the WTO.

I argue that firms protect their interests through the dispute settlement process by contributing to the litigation costs of a WTO dispute, while governments use firm contributions to screen potential WTO complaints. When a government is unwilling to pursue a case due to high litigation costs or their belief that the case is weak, firms can step in to fill the gap between expected costs and expected profits and to signal the strength of the case. Of course, governments still retain control over the gatekeeping process, so if the diplomatic externalities of the case are too high, the government may choose not to bring the case, which is a key distinction between the legal procedures of the WTO and transnational dispute settlement mechanisms.

Firms have an informational advantage throughout the litigation process, given their position in perceiving and analyzing the trade barriers they face. The unique positions of firms can best be illustrated by considering their role in the three phases of litigation known as “naming, blaming and claiming” (Felstiner, Abel, and Sarat 1981). The naming phase involves identifying an injury to one’s trading prospects (Shaffer 2006). The difficulty and cost of naming varies depending on the type of potential violation. For example, when antidumping duties are implemented the country imposing them must notify the exporting
firms making them particularly easy to identify, whereas the provision of subsidies that lower the cost to a competitor and reduce a company’s market share will be much more difficult and costly to identify (Bown 2009). Whether the cost is high or low, the private industry has the greatest incentive and ability to identify an economic injury. Through regular business practices, firms will be the first to experience the negative effects of WTO inconsistent measures, which means the costs of naming for private firms are relatively low when compared to the potential costs to the government.

The “blaming” phase of a dispute determines who is responsible for the injury identified in the first phase (Shaffer 2006). Once the injury is perceived, the blaming phase can be relatively straightforward. If the lost profits are due to a trade disruption with a specific trading partner or to a flood of imports from a specific country, minimal costs should be associated with identifying who is to blame.

Once the naming and blaming have been completed, the most expensive and complex phase of dispute settlement begins - “claiming.” This final phase consists of developing and pursuing a legal claim through the WTO (Shaffer 2006). Expenditures incurred during this phase include, but are not limited to, research costs, legal fees, administrative outputs, and travel expenses. A USTR official interviewed for this project estimates that half to three-quarters of the litigation expenses are devoted to the fact finding portion of claiming (USTR Official 2009). During this phase, private firms will quantify the value of lost revenue from trade, build the case connecting their losses to the barrier in question, and then work with government to formalize the complaint through the WTO dispute settlement process.

The comparative advantage of firms in naming, blaming, and claiming enables them to act as fire alarms, in the sense developed by McCubbins and Schwartz (1984), identifying and signaling the existence of harmful trade violations to their government. On the other hand, government efforts to act as “police patrols” are a relatively inefficient mechanism when compared to private firms. This creates an environment where the private firms have
an information advantage, uniquely positioning firms to monitor and signal cases to the
government in an effort to protect their interests.

An example of this type of public-private relationship occurred in a WTO dispute over
genetically modified foods between the European Community and the United States, DS291.
In this case the United States, along with Argentina and Canada, requested the formation
of a dispute settlement panel on August 7, 2003 (World Trade Organization 2012a). Prior to
the initiation of consultations and the formal request for a panel, Monsanto, a producer of
genetically modified foods, which had 15 products that had allegedly been adversely affected
by the European Community’s actions (World Trade Organization 2012b), directly engaged
the US government in an effort to ensure the case was brought.

According to interviews with a USTR official, when deciding whether to initiate the case,
the CEOs from the affected companies met with USTR officials and agreed to support the
litigation effort (USTR Official 2009). In order to convince the government to bring the case,
the firms had to fund and complete a “laundry list” of fact-finding and litigation assignments
(USTR Official 2009). In this case, a relationship was built where the private firms showed
their beliefs about the value and strength of the case by taking on a substantial portion of
the fact-finding responsibilities and expenses. In response to the signals of the firms, the
USTR moved forward with the case with greater confidence in the strength of the case and
at a drastically reduced cost.⁶

From the perspective of the government, private party contributions are also important
for relaxing the government’s budget constraint. For example, the USTR is responsible for
initiating WTO complaints for the US, but their total budget is only about $47.5 million
annually (Cook 2013). Within their budget, the executive’s top priorities are negotiating
trade agreements – not litigating existing agreements (USTR 2014). This creates a situation

⁶Firms’ litigation contributions may also increase the legal strength of the case by providing improved
argumentation and additional evidence. In this paper, I do not evaluate the connection between contributions
and legal strength, which means the model provides a conservative approach to evaluating the importance
of firms’ contributions. I thank an anonymous reviewer for emphasizing this point.
where, as the USTR’s top litigator noted, budget concerns limit the ability to initiate new legal complaints and seek enforcement of existing trade agreements \(^{7}\) (World Trade Online 2013). Private firms’ contributions can thus make a significant difference in which cases are likely to be brought. Government officials charged with seeking enforcement of trade agreements can increase their chance of success and their effective litigation budgets by screening cases based on firm contributions.

The firms’ contributions also play an informational role as a signal of the strength of the case, which is a key factor in determining when the government is willing to challenge potential WTO violations. Examining the European context, Chalmers (2013) notes that the “currency of lobbying in the European Union (EU) is information.” Firms’ litigation contribution functions as a signal of the firms’ private information about the strength of the case, allowing the government to only pursue the strongest complaints. This is particularly important given governments’ risk aversion when initiating WTO disputes. Two primary factors contribute to governments’ heightened risk aversion, compared to firms. First, the government officials responsible for selecting cases face constrained budgets, and must choose from a broader set of potential cases to only initiate a select few. With this in mind, officials seek to pursue cases where they are most likely to use their resources effectively and be perceived as selecting the best cases. A USTR official highlighted the importance of screening out weak cases, noting that they seek “slam dunk” cases, while a member of the European Commission noted that the “strength of the legal issue” is of primary importance (USTR Official 2009; European Commission Official 2009). These statements reflect the unique risk to the government of pursuing a weak case. While firms also face resource constraints, each firm has a smaller set of potential disputes to choose from, and pursuing the firms’

\(^{7}\) As in the case of DS291, the government can reduce its legal expenses when firms contribute to litigation costs by taking on responsibilities that might otherwise be considered the role of the state. Rather than having the state use government attorneys or publicly financed representation to prepare case-materials, firms can engage in the naming, blaming, and claiming process by conducting research, preparing legal briefs, and litigating the case on behalf of the state.
strongest case may still be somewhat of a gamble, whereas the government officials have the opportunity to select a pool of strong cases, and are best off choosing only the strongest. Additionally, when a government pursues and loses a WTO complaint, they not only face the losses from the dispute in question, but they also face a changed legal landscape where the issue in question is given a green light by the WTO. This raises the cost of losing a complaint, because there may be potentially far reaching externalities from the adoption of similar policies by other trade partners. For example, if the US were to file and lose a complaint against China regarding currency manipulation, not only would China be able to continue their policies, but other countries would now be able to adopt similar policies without fear of legal challenges (Davis 2012b, 165-168). Due to the risks associated with losing a complaint, governments place significant weight on the strength of cases when evaluating whether to challenge potential WTO violations.

1.4 The Model

While firms have an incentive to signal the strength and value of cases to their government, the government and firms’ preferences are not necessarily perfectly aligned. As previously noted, firms may be more willing to pursue legally weak cases, which means the government must screen out potentially insincere signals from firms. The formalization of the signaling logic is particularly useful in identifying a threshold for firms’ litigation contributions, such that the signal is sufficiently costly so the government can infer the credibility of the message. Without this threshold, firms could easily bluff and attempt to convince the government to pursue poor cases. Unlike traditional informational lobbying models, the litigation contribution model incorporates the added effect of mitigating the government’s resource constraint, which expands the set of cases that are initiated in equilibrium.

The structure of the model is designed to capture key dynamics of the WTO case selection and litigation process. I begin with a basic form of the model with just two players, Firm
and Government. The subscripts $F$ and $G$ are used to identify the actions of each respective player. The information structure captures the existence of asymmetric information between the government and firm. Both players are assumed to have some prior knowledge of the expected strength for a potential case, but the firm receives more accurate information than the government. Because firms are engaged in trading and experience the direct cost of any trade barrier, firms are able to identify potential trade violations and assess the potential strength with greater ease and accuracy than the government. This asymmetry is captured in the model when the firm receives a message about the strength of the case, $m \in \{s, w\}$, but the government does not. The firm’s private information means it has more accurate knowledge of the probability of winning a particular case than the government.

The setup of the model also captures the formal rules of the WTO. The WTO Dispute Settlement Understanding limits dispute initiation to governments, giving them the final (and only) decision about whether to file a complaint with the WTO. The model captures this structure by maintaining the status quo trade barrier if the government does not initiate a case, and introduces a possibility of its removal only when the government initiates. The firm decides whether to contribute ($L_F > 0$) or not ($L_F = 0$). When the firm decides to contribute, it selects a contribution level to the litigation effort, which captures the firm’s ability to contribute time and resources to the naming, blaming, and claiming of the litigation process.

The model begins when the players are presented with a potential WTO dispute. The potential case is exogenously given, as is the total cost of litigation, the probability the case is won, and the value of winning the case. The trade value of winning a case is defined as the benefits from trade with the trade barrier removed minus the benefits from trade with the trade barrier in place, which is written as $\tau_j(0) - \tau_j(1)$, where $j \in \{F, G\}$. The value of a case will depend on the level of distortion caused by the trade barrier and magnitude of

---

8The two-player version of the model focuses on asymmetric information about the strength of the case; however, a similar logic holds if the uncertainty is about the value of the case.
the affected trade flow, but for simplicity the payoffs can be normalized such that $\tau_j(1) = 0$ and $\tau_j(0) = 1$, so the trade gains for both players are 1 if the case is won. To allow for the possibility of externalities to the government, the model includes an externality term, $E_G$. Externalities for a case can range from non-trade retaliation, such as restricted foreign aid, to increasing political support for the government from appearing to stand strong with domestic industry. The total litigation cost for the case is $L$. The total litigation cost represents the expected cost of bringing a case, including all phases of naming, blaming, and claiming.

The game begins with nature determining whether a particular case is strong or weak, $\theta_S$ or $\theta_W$. A strong case is defined such that the probability of winning the case is uniformly distributed between 0.5 and 1.0. A weak case is defined such that the probability of winning the case is uniformly distributed between 0.0 and 0.5. The players’ priors over the strength of the case are that with probability $P$ the case is strong and with probability $1 - P$ the case is weak.

After the firm receives a message about the strength of the case, which is unknown by the government, the firm decides to contribute or not. If the firm contributes it pays a cost, $L_F > 0$. The firm selects the exact cost it pays, which is deducted from the total cost of litigation. Once the firm has acted, the government is faced with the decision whether to initiate a WTO complaint or not, $I_G$ or $\neg I_G$. If the government initiates it pays $L_G = L - L_F$, and has an expected payoff of $EU_G(I_G) = \hat{\theta}_{W,S}(\tau_G(0)) - L_G + E_G$. If the government does not initiate it has an expected payoff of $EU_G(\neg I_G) = \tau_G(1)$. The payoffs capture two of the most important elements of the case selection process, the probability the case is won and the value of the case. Using these payoffs as the selection criteria for WTO case initiation is supported by the case selection practices of trade officials. For example, a former General Counsel with the USTR noted in an interview with the author that economic considerations are the most important factor in deciding if a complaint is initiated ([USTR General Counsel](2009)). Another USTR official stated in an interview that the magnitude of the expected profit is
important, but the probability of success is the most important factor \cite{USTROfficial2009}. The model thus captures two of the key elements of the WTO dispute initiation process and provides useful insights into the dynamics of international trade law enforcement. The extensive form version of the game is shown in Figure 1.4.

An obvious result of the model is that a case will not be initiated if \( L > \hat{\theta}_S(\tau_G(0)) + E_G + \hat{\theta}_S(\tau_F(0)) \). This result simply states that if the litigation cost for a case is greater than the combined expected payoff to the government and firm, then the case is not profitable to pursue. A further general result of the model is that whenever \( L < \hat{\theta}_W(\tau_G(0)) + E_G \) the firm will choose not to contribute, because the government will initiate the case without a contribution from the firm. In this situation, \( L_G = L \) and \( EU_G(I_G) > 0 \). This means that the litigation cost of the case is low enough relative to the expected payoff that it is beneficial for the government to unilaterally initiate the case. Although rare, these types of cases would likely be brought when the precedent value of a case is high, which occurred in some of the early intellectual property rights disputes \cite{USTROfficial2009}.

The above results are affirmed in the Perfect Bayesian Equilibrium of the model. The most interesting results of the model are from the set of cases where the government would be unwilling to initiate the case without a litigation contribution from the firm. The first set of such cases are those where the expected profit to the government is less than the total litigation cost. In a unitary actor model, these cases would be viewed as unprofitable, however the equilibrium result shows that the firm’s litigation contribution can alter the expected payoffs to the state, making such cases profitable to the government and increasing the universe of profitable cases. A second, and potentially overlapping, group of cases are those where the government’s prior belief regarding the strength of a case is sufficiently low that the government does not believe case initiation is profitable. In this group of cases,

\footnote{It has become widely accepted that the \textit{de facto} importance of precedent can be quite high in WTO disputes \cite{Bhala1999, BuschandPelc2010, Pelc2014}.}

\footnote{The proof is provided in the appendix, section A-2.}
if the firm knows that the case is strong, it can credibly signal the strength of case to the government, thus altering the expected payoffs of the government and motivating the government to initiate the case. When \( L_F \geq L - \hat{\theta}_S(\tau_G(0)) - E_G \) and \( L_F \geq \hat{\theta}_W(\tau_F(0)) \)
the Firm has contributed a sufficient amount, such that the government now believes its expected payoff from case initiation is greater than or equal to zero and the government initiates the case. For simplicity, I will refer to this contribution threshold for the firm as $L^*_F$.

In order for the firm’s signal to be credible, the equilibrium condition requires that the litigation contribution of the firm, $L^*_F$, must be greater than the firm’s expected profit from a weak case. The litigation contribution threshold means that the government does not update its beliefs about the strength of the case when the firm contributes less than $L^*_F$. This means there exists a unique separating equilibrium where firms will only contribute $L^*_F$ when they know a case is strong.\footnote{\textsuperscript{11} Proof of the equilibrium is provided in the appendix, section A-2. Proof of uniqueness is in the appendix, section A-3.} This separating equilibrium helps explain the extremely high success rate of WTO complainants, given that governments are able to screen out potentially week cases when working with private firms during the litigation process.\footnote{\textsuperscript{12} Although some models of judicial case selection would suggest that defendants would anticipate this selection process and avoid going to trial when cases are strong, Davis (2012b, 88) explains why many members of the WTO “stonewall” trade settlements and instead go to trial at the WTO.}

The equilibrium contribution levels for both the firm and government (for $P = .5$ and $E_G = 0$) are shown in Figure 1.4. The figure shows that if the total litigation cost is low enough ($L \leq .5$), then the firm pools on contributing nothing and the government pays the full amount and initiates on its own. In the next portion of the parameter space ($.5 < L < .75$), the firm pools on contributing $L - .5$, which is just enough to make the government initiate the case, but does not convey a credible signal and thus the government does not update its beliefs about the strength of the case. In the next portion of the parameter space ($.75 \leq L \leq 1.5$), the firm strategies fully separate, with contributions equal to zero when the case is weak (right panel) and contributions equal to $L^*_F$ if the case is strong (left panel). In this range of potential disputes, the firm’s signal is informative and allows the government to only pursue cases that are strong. Lastly, once cases become prohibitively
costly \( (L > 1.5) \), the firm again pools on contributing nothing and the government does not initiate. The figure illustrates that for any given set of parameters there is a unique contribution for both the firm and government.\textsuperscript{13}

A further comparative static of the model is that, all else equal, a case will be more likely to be initiated when the distortion caused by a particular trade barrier is greater. A higher level of distortion means that a country will be forgoing relatively more trade, which increases the value of \( \tau_j(0) - \tau_j(1) \). Distortion also acts as a proxy for legal strength, given that proving economic harm can be an important facet of achieving compensation and securing a legal victory, and is indeed required for Article XXIII nullification or impairment complaints. Distortion impacts the expected profit and strength of the case, which means trade barriers with higher distortion should be contested in the WTO with a higher probability than similar barriers with lower levels of distortion.

The comparative statics of the model with the Government and Firm provide a starting point for understanding WTO case initiation, but I now consider the incentives for an industry with multiple firms. Evaluating the likelihood of firm contributions within an industry with multiple firms allows me to examine how heterogeneous preferences across firms affects the case selection process. I begin by allowing for multiple firms within an industry affected by a potential trade barrier to have heterogeneous preferences with regard to the potential dispute. While firms may still have better knowledge about the strength of a case, I now examine how uncertainty over the heterogeneous valuations of the firms affect the likelihood they contribute to the litigation process.\textsuperscript{14} This is formalized by the existence of \( N \geq 2 \) firms in an industry where each firm \( i \in N \) values bringing the case with a utility of \( V_i \), where \( V_i = \tau(0) - \tau(1) \) which is independently drawn from a continuous distribution \( F \). \( V_i \) is

\textsuperscript{13}The proof is provided in the appendix, section A-2.
\textsuperscript{14}Since the expected payoffs to firms are a function of both the strength of a case and the firms' valuations, all else equal, firms are still more likely to contribute when the case is strong instead of weak. However, taking strength of case as a given allows me to examine how the structure of industries and firms valuations for disputes affects firm participation in the WTO dispute process.
Figure 1.2: Firm and Government Equilibrium Contributions

Figure 4 plots the unique equilibrium contributions for the firm and government over a range of total litigation costs ($L$) and a set value for $P$ and $E_G$. Comparing the left and right panels of the figure shows that for certain ranges in the parameter space firm strategies pool, but in the middle range of the parameter space ($0.75 \leq L \leq 1.5$) the strategies fully separate based on the strength of the case. Firms thus have a unique strategy profile with a unique contribution level for each combination of parameters, resulting in a separating equilibrium.

private information, known only by firm $i$, although firms are aware of the distribution from which other firms values are drawn. The model also allows for litigation contribution levels to vary across firms.

In the model with multiple firms, the Government receives a contribution from the industry, which is the sum of the contributions from all firms within the industry, which is still $L_F$. Given the separating equilibrium from the earlier analysis, when firms contribute a combined $L_F^*$ the government will choose to initiate the WTO dispute. As in the previous analysis, the choice to bring the case by the government is dichotomous; it either does or does not initiate the case. With these constraints in mind, the model of firm contributions to the litigation process perfectly resembles a contribution game where private actors with
incomplete information engage in a game to provide a discrete public good. In this case the
public good can be thought of as the initiation of the case, where the good is the benefit from
the case that accrues to the firms within a given industry. Of course, not all firms within an
industry will benefit equally from a given trade dispute, which is why firm level valuations
for the case are modeled as heterogeneous. A more complete discussion of the model with
multiple firms is provided in the appendix, section A-4.

Consistent with the model with just one firm, firms’ decisions are modeled as taking place
simultaneously in a one-shot contribution game. The contribution game specifies that all
costs that are paid by firms to the litigation expenses are not refundable, which is consistent
with the practice of the case selection process. This game has been analyzed by Menezes,
Information.”

In the most simplistic version of the model, I allow the cost of contributing the good to
be low enough such that a single firm can initiate the case. In this situation a single firm will
contribute $L_F^*$ and the good is provided, which means the Government initiates the case.
Although multiple firms could contribute to the cost of bringing the case, the symmetric
equilibrium exists where a firm with a sufficiently high payoff will provide the good on its
own (Menezes, Monteiro, and Temimi 2001, 499).

The first notable comparative static to emerge from the model with incomplete informa-
tion and heterogeneous firms and contributions is that industries with dominant firms will
be more likely to initiate cases, since it is more likely that a dominant firm will be able to
afford to pay $L_F^*$. This finding hinges on the fact that an industry where a single firm has
a relatively high expected payoff from a WTO case, such that $L_F^*$ is less than or equal to
$V_i$, has a strictly greater probability of contributing to the litigation cost of a dispute than
an industry where no single firm has an incentive to pay $L_F^*$, in which case the probability
that a case is initiated is strictly less than one (unless the case is initiated unilaterally by the
Government). Furthermore, a dominant firm will also be most likely to have the capacity to pay \( L^*_F \). All else equal, in industries where dominant firms have relatively high expected payoffs and capacity to pay the litigation contribution, the probability that there exists a single firm willing to pay \( L^*_F \) will be greater than in an otherwise identical industry. Thus, we should expect to see more case initiation in industries with high capacity dominant firms.

Next I consider the contribution game when no single firm can afford to pay \( L^*_F \), and I find that a coordination problem exists that eventually becomes great enough that a symmetric equilibrium resulting in case initiation is no longer possible. For a wide range of costs of a public good, the coordination problem prohibits provision of the good (Menezes, Monteiro, and Temimi 2001, 496). Of particular importance is the finding that if the cost of the public good is slightly above the aggregate mean of the valuations then the unique equilibrium of the contribution game is for each player to contribute zero no matter what its value is (Menezes, Monteiro, and Temimi 2001, 502). This finding implies that even when an industry as a whole may stand to benefit from the initiation of a WTO dispute, if no single firm can afford to pay the necessary litigation cost to motivate the government to file and the average valuation by all firms within the industry is low enough, the case will not be initiated. \(^{15}\)

From this, a second comparative static emerges — as the mean value and capacity for the industry increases, case initiation becomes more likely, since there is a greater chance that the mean value and capacity for the industry will exceed the cost of litigation, which makes it more likely firms will contribute to the litigation process.

### 1.5 Evidence of Firm Signaling and Contributions

I now test the implications of the model using firm-level data gathered from Compustat in conjunction with the Foreign Trade Barrier Dataset (FTBD), which allows me to test

\(^{15}\)It can also be shown that the probability that the good is provided in this situation is strictly less than one, even when the mean contribution does not exceed the prohibitive threshold and when the outcome would be efficient (Menezes, Monteiro, and Temimi 2001, 496).
the effect of firms’ litigation capacity, the level of trade barrier distortion, and competing theories on the probability of dispute initiation from a set of potential WTO cases. The FTBD is comprised of potential disputes, which are defined as harmful trade barriers to US exports identified in the National Trade Estimate (NTE) annual reports (Davis 2012a). This dataset has a unique advantage over previous datasets that examined exclusively antidumping measures or self-reported trade barriers. Unlike previous datasets, the FTBD encompasses non-tariff barriers and regulations that affect a range of industries, investment policies, and trade standards as perceived by the “victim,” the US, between 1995 - 2004. The data allow me to test the model’s implications within a subset of potential trade barriers that have met a minimum threshold to be recognized by the government. The data are restricted to trade barriers against the US, which has the advantage of holding the initiating country constant, which controls for a multitude of potential covariates, but also means there is room for future extensions with new data covering a broader set of countries.

The unit of analysis is the trade barrier, with an observation included for every year the NTE mentions the barrier in their report. I begin by testing the influence of distortion caused by a trade barrier on the probability that the barrier is challenged in the WTO. The Distortion variable for each trade barrier is coded as an indicator variable that identifies cases with significant market closure resulting from a ban, quota, or increase of tariff/duty of more than 10 percent, standards or rules of origin that create a de facto ban on imports, violation of intellectual property rights, or subsidies to competitors (Davis 2012b). The expectation for distortion is positive, as the variable directly increases the payoff from the case and the expected legal strength, as analyzed in the model.

To test the the effect of industry-specific variables, the industry affected by the trade barrier is coded at the level of the ISIC3 4 digit classification. The model predicts that

---

16 Trade barriers may be reported to the USTR, and thus enter the NTE reports, via numerous methods, including a telephone hotline or online reporting. This means there is a relatively low threshold for barriers to enter the dataset; however, if some low-value barriers are left out, this would attenuate the results and is thus not a major concern for this paper. I thank an anonymous reviewer for raising this point.
across industries, those industries with dominant firms with the capacity to pay the litigation costs and with a sufficiently high value for the dispute are most likely to initiate cases. To examine the capacity of dominant firms within industries, I compiled firm-level data using the Compustat database. For each industry, I test the effect of *Dominant Firm Capacity*, measured as the log of the earnings in a given year for the top earning firm in the industry. I also test the *Average Firm Capacity* for each industry, using the average earnings for each industry in a given year. Together, these measures allow me to examine two of the comparative statics that emerge from the model with heterogenous firms and incomplete information.

To control for other case factors and to test competing hypotheses I include a range of controls. First, I examine whether progress has been made in negotiating the removal of the trade barrier. *Progress* is coded on a four point scale indicating the level of progress toward resolving the disputed trade barrier (Davis 2012a). Because a WTO dispute is a costly means of removing a trade barrier, I expect that if significant progress is being made through other means a WTO complaint will be less likely. I also control for the length of time, *Duration*, the trade barrier has been reported in the NTE. The expected sign for duration is negative, as barriers that have been constant over time are less likely to be challenged than new barriers that suddenly disrupt trade flows. Because the universe of cases is limited to the US, I also control for whether there is an active Section 301 petition (*Active 301*). An active 301 petition requires government attention and is expected to have a positive influence on the probability a case is initiated.

To test competing hypotheses, I also test the influence of industry size and political contributions, directed exports to the trade partner, and country specific effects. The models presented in this paper suggest that firms with the capacity to contribute to the litigation process play a critical role in the case selection of disputes by signaling information about

---

17 The specific earnings are defined as “retained earnings”, which are the cumulative earnings of the company less total dividend distributions to shareholders.
the strength of a case. A plausible counter argument is that well-organized industries could buy litigation through traditional lobbying such as campaign contributions and electoral support. I test this competing theory by evaluating the effect of industries’ political contributions to political parties and politicians. This variable is coded as the log of total political contributions in constant year 2000 dollars, as reported by the Center for Responsive Politics (Davis 2012a). To evaluate whether industry size is driving the results, I also test the value of production of the industry, measured as the log of its total production (Davis 2012a). If direct political influence or the production value of the industry is the driving force, then political contributions and production value should have a positive effect on whether a case is initiated. I also test the theory that relative economic power plays a significant role in dispute initiation, which is derived from the self-enforcing nature of the WTO dispute settlement system. Although a measure of the complainant’s economic relationship with the defendant will have relatively little variation in my sample (compared to a cross-national study of complainants), I test this proposition with the log of the annual exports from the US to the trade partner imposing the barrier.

Because the dependent variable of interest is a dichotomous decision whether or not to file a WTO complaint for a particular trade barrier in a given year, I use a logistic regression to analyze the data. I employ a random effects model with groups defined at the ISIC3 4 digit level, which addresses the fact that a number of the variables, including key treatment variables, occur at the industry level. This selected model identifies intercepts by industry, while allowing for the effects of the key variables of interest to be analyzed across the dataset. I first test the influence of dominant firm capacity and distortion and the results are reported in Table 1.1

The results of the random effects models illustrate the positive and significant influence Distortion and Dominant Firm Capacity have on the probability a trade barrier is challenged

---

The results are robust to grouping on trade barrier as Davis (2012) does.
in the WTO. Model 1 analyzes the results in a basic form, while Model 2 controls for country specific effects among some of the primary trading partners of the US. The results demonstrate that the variables of interest are not sensitive to country level controls and the additional controls of Progress and Duration have the expected sign. Robustness checks were run with Mexico, Korea, or non-OECD countries as the base-group (results not reported here), and the results were consistent across specifications.

Model 3 and Model 4 introduce variables testing competing theories of case selection, in addition to controlling for an active Section 301 petition in Model 4. Model 3 tests the impact of (the log of) total exports of the US to the trade partner, which does not have a significant effect. I also examine the possibility that the relative value of the industry might be accounting for the significance of the dominant firm’s capacity, which would occur if the the presence of a high capacity dominant firm was highly correlated with the production value of the industry. Model 3 controls for the log of the industry’s production value and demonstrates that even when controlling for the industry’s size, dominant firm capacity remains a significant factor in determining dispute initiation. Lastly, I examine whether other forms of political influence, such as direct lobbying have a significant effect. Model 4 shows that political contributions are not statistically significant. The Section 301 petition has the anticipated effect and is statistically significant in the sparser model, but not so in the full specification in Model 5.

\[\text{It is also possible that certain industries are more or less likely to be involved in disputes, due to the type of business they do or other traits constant to the industry. To address this concern, I replicate the results with fixed effects at the industry level. The results are consistent with those reported in Table 1.1 and are discussed in the appendix, section A-5.}\]

\[\text{The results from Model 5 are also consistent when controlling for the value of exports from the affected industry (instead of US Exports to Trade Partner), although data limitations result in more than a third of the observations being dropped, which results in some key variables loosing statistical significance.}\]

\[\text{Due to data availability the number of observations fluctuates across models. To examine whether changes in significance are driven by changes in the sample, particularly for an active Section 301 petition, all results are replicated using the same sample of 1407 observations. The results are consistent with those reported in Table 1.1 and are displayed in the appendix, section A-5.}\]
<table>
<thead>
<tr>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
<th>Model 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dominant Firm Capacity</td>
<td>0.286**</td>
<td>0.312**</td>
<td>0.304*</td>
<td>0.265*</td>
</tr>
<tr>
<td></td>
<td>(0.142)</td>
<td>(0.150)</td>
<td>(0.154)</td>
<td>(0.157)</td>
</tr>
<tr>
<td>Trade Barrier Distortion</td>
<td>2.361***</td>
<td>2.427***</td>
<td>2.257***</td>
<td>2.179***</td>
</tr>
<tr>
<td></td>
<td>(0.766)</td>
<td>(0.784)</td>
<td>(0.776)</td>
<td>(0.771)</td>
</tr>
<tr>
<td>Negotiation Progress</td>
<td>−1.235***</td>
<td>−1.168***</td>
<td>−0.922*</td>
<td>−1.280***</td>
</tr>
<tr>
<td></td>
<td>(0.442)</td>
<td>(0.442)</td>
<td>(0.467)</td>
<td>(0.444)</td>
</tr>
<tr>
<td>Trade Barrier Duration</td>
<td>−0.262**</td>
<td>−0.252*</td>
<td>−0.243*</td>
<td>−0.250*</td>
</tr>
<tr>
<td></td>
<td>(0.131)</td>
<td>(0.134)</td>
<td>(0.136)</td>
<td>(0.129)</td>
</tr>
<tr>
<td>EU</td>
<td>1.108</td>
<td>1.464</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.111)</td>
<td>(1.276)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Japan</td>
<td>0.870</td>
<td>0.306</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.198)</td>
<td>(1.547)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mexico</td>
<td>1.649</td>
<td>0.284</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.278)</td>
<td>(1.921)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Korea</td>
<td>0.636</td>
<td>−3.513</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.262)</td>
<td>(2.819)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NonOECD</td>
<td>0.092</td>
<td>−5.167</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.180)</td>
<td>(3.969)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>US Exports to Trade Partner</td>
<td>0.144</td>
<td></td>
<td>−2.016</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.235)</td>
<td></td>
<td>(1.294)</td>
<td></td>
</tr>
<tr>
<td>Industry Production</td>
<td>−0.059</td>
<td></td>
<td>0.310</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.268)</td>
<td></td>
<td>(0.444)</td>
<td></td>
</tr>
<tr>
<td>Political Contributions</td>
<td>−0.232</td>
<td>−0.255</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.196)</td>
<td>(0.328)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Active 301</td>
<td>1.620**</td>
<td>1.468</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.648)</td>
<td>(0.984)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>−7.999***</td>
<td>−8.975***</td>
<td>−10.894</td>
<td>−4.099</td>
</tr>
<tr>
<td></td>
<td>(1.536)</td>
<td>(1.997)</td>
<td>(7.158)</td>
<td>(3.213)</td>
</tr>
</tbody>
</table>

Random effect models calculated using xtmelogit with STATA10. Random intercepts calculated for groups at the industry level, defined as the ISIC3 4 digit industry. Canada is the omitted comparison. P-values are calculated using a two-tailed test.

To evaluate the substantive significance of the findings, I estimate the predicted probabilities of filing a WTO complaint given varying levels of dominant firm capacity, trade barrier distortion, and progress while holding the remaining variables at their means or a value of 1 for dichotomous variables. I use Model 5 from Table 1.1 which controls for an array of competing variables and country effects, and thus is the preferred specification. From the predicted probabilities, I calculate the change in the probability of dispute initiation for a

---

*p < .1, **p < .05, ***p < .01

---

21 The defendant country is Mexico, so other country dummies and the non-OECD countries are set to 0.
shift from one standard deviation below the mean to one standard deviation above the mean in significant variables, which are reported in the top panel of Figure 1.5.

The predicted probabilities of filing a complaint with dominant firm capacity one standard deviation below the mean, when the hypothetical defendant is Mexico, is 0.20. The same probability with the dominant firm’s capacity one standard deviation above the mean is 0.40. In practice, this change is approximately the effect of changing from an industry manufacturing games and toys to an industry manufacturing pharmaceutical products. Similarly, the predicted probability for a case with distortion equal to 0 is 0.07, as opposed to 0.28 with high distortion. Clearly, both the capacity of the dominant firm and distortion have a significant effect on the likelihood a trade barrier is challenged.

In the same progression as Table 1.1, I test the impact of average firm capacity with the results reported in Table 1.2. The impact of average firm capacity is robust to country specific effects and the full range of controls for competing theories. Using the full specification from Model 5, I examine the substantive influence of average firm capacity on dispute initiation, with the results displayed in the lower panel of Figure 1.5. The change in average firm capacity increases the probability of case initiation from 0.20 to 0.36, highlighting the substantive impact of average firm capacity.
Figure 1.3: Change in Predicted Probability Plots

Effect of Significant Variables
on the Probability of Dispute Initiation

Change in predicted probability is calculated from Model 5 of Table 1

Change in predicted probability is calculated from Model 5 of Table 2. Estimates and 90 percent confidence intervals are calculated using a quasi-bayesian simulation that samples 2000 times from a distribution based on the coefficients and variance. Changes in predicted probabilities represents a shift from one standard deviation below the mean to one standard deviation above the mean of the variable, or a shift from 0 to 1 for distortion. All other variables are held at their mean or one for indicator variables, and the hypothetical defendent is Mexico.
Table 1.2: Random Effects Logistic Regression of WTO Dispute Complaints

<table>
<thead>
<tr>
<th></th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
<th>Model 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average Firm Capacity</td>
<td>0.0007**</td>
<td>0.0008**</td>
<td>0.0009***</td>
<td>0.0006*</td>
<td>0.0009**</td>
</tr>
<tr>
<td></td>
<td>(0.0003)</td>
<td>(0.0003)</td>
<td>(0.0003)</td>
<td>(0.0003)</td>
<td>(0.0004)</td>
</tr>
<tr>
<td>Trade Barrier Distortion</td>
<td>2.177***</td>
<td>2.166**</td>
<td>2.070***</td>
<td>2.035***</td>
<td>1.987***</td>
</tr>
<tr>
<td></td>
<td>(0.766)</td>
<td>(0.775)</td>
<td>(0.774)</td>
<td>(0.767)</td>
<td>(0.792)</td>
</tr>
<tr>
<td>Negotiation Progress</td>
<td>−1.139**</td>
<td>−1.021**</td>
<td>−0.879*</td>
<td>−1.256***</td>
<td>−0.964**</td>
</tr>
<tr>
<td></td>
<td>(0.442)</td>
<td>(0.442)</td>
<td>(0.467)</td>
<td>(0.441)</td>
<td>(0.470)</td>
</tr>
<tr>
<td>Trade Barrier Duration</td>
<td>−0.269**</td>
<td>−0.262*</td>
<td>−0.252*</td>
<td>−0.254*</td>
<td>−0.144</td>
</tr>
<tr>
<td></td>
<td>(0.130)</td>
<td>(0.134)</td>
<td>(0.135)</td>
<td>(0.131)</td>
<td>(0.156)</td>
</tr>
<tr>
<td>EU</td>
<td>0.905</td>
<td>1.213</td>
<td>1.213</td>
<td>1.213</td>
<td>1.213</td>
</tr>
<tr>
<td></td>
<td>(1.097)</td>
<td>(1.273)</td>
<td>(1.273)</td>
<td>(1.273)</td>
<td>(1.273)</td>
</tr>
<tr>
<td>Japan</td>
<td>0.483</td>
<td>−0.109</td>
<td>−0.109</td>
<td>−0.109</td>
<td>−0.109</td>
</tr>
<tr>
<td></td>
<td>(1.192)</td>
<td>(1.573)</td>
<td>(1.573)</td>
<td>(1.573)</td>
<td>(1.573)</td>
</tr>
<tr>
<td>Mexico</td>
<td>1.397</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>(1.278)</td>
<td>(1.952)</td>
<td>(1.952)</td>
<td>(1.952)</td>
<td>(1.952)</td>
</tr>
<tr>
<td>Korea</td>
<td>0.165</td>
<td>−4.164</td>
<td>−4.164</td>
<td>−4.164</td>
<td>−4.164</td>
</tr>
<tr>
<td></td>
<td>(1.262)</td>
<td>(2.929)</td>
<td>(2.929)</td>
<td>(2.929)</td>
<td>(2.929)</td>
</tr>
<tr>
<td>NonOECD</td>
<td>−0.395</td>
<td>−5.587</td>
<td>−5.587</td>
<td>−5.587</td>
<td>−5.587</td>
</tr>
<tr>
<td></td>
<td>(1.200)</td>
<td>(4.043)</td>
<td>(4.043)</td>
<td>(4.043)</td>
<td>(4.043)</td>
</tr>
<tr>
<td>US Exports to Trade Partner</td>
<td>0.253</td>
<td>−2.037</td>
<td>−2.037</td>
<td>−2.037</td>
<td>−2.037</td>
</tr>
<tr>
<td></td>
<td>(0.259)</td>
<td>(1.305)</td>
<td>(1.305)</td>
<td>(1.305)</td>
<td>(1.305)</td>
</tr>
<tr>
<td>Industry Production</td>
<td>−0.137</td>
<td>0.070</td>
<td>0.070</td>
<td>0.070</td>
<td>0.070</td>
</tr>
<tr>
<td></td>
<td>(0.258)</td>
<td>(0.443)</td>
<td>(0.443)</td>
<td>(0.443)</td>
<td>(0.443)</td>
</tr>
<tr>
<td>Political Contributions</td>
<td>−0.221</td>
<td>−0.114</td>
<td>−0.114</td>
<td>−0.114</td>
<td>−0.114</td>
</tr>
<tr>
<td></td>
<td>(0.192)</td>
<td>(0.343)</td>
<td>(0.343)</td>
<td>(0.343)</td>
<td>(0.343)</td>
</tr>
<tr>
<td>Active 301</td>
<td>1.599**</td>
<td>1.360</td>
<td>1.360</td>
<td>1.360</td>
<td>1.360</td>
</tr>
<tr>
<td></td>
<td>(0.663)</td>
<td>(1.015)</td>
<td>(1.015)</td>
<td>(1.015)</td>
<td>(1.015)</td>
</tr>
<tr>
<td>Constant</td>
<td>−5.697***</td>
<td>−6.126***</td>
<td>−10.23</td>
<td>−2.162</td>
<td>47.409</td>
</tr>
<tr>
<td></td>
<td>(0.756)</td>
<td>(1.234)</td>
<td>(7.534)</td>
<td>(3.176)</td>
<td>(33.853)</td>
</tr>
<tr>
<td>N</td>
<td>1635</td>
<td>1635</td>
<td>1418</td>
<td>1624</td>
<td>1407</td>
</tr>
</tbody>
</table>

*p < .1, **p < .05, ***p < .01

Random effect models calculated using xtmelogit with STATA10. Random intercepts calculated for groups at the industry level, defined as the ISIC3 4 digit industry. Canada is the omitted comparison. P-values are calculated using a two-tailed test.

The empirical results provide valuable support for the theoretical model of this paper, but they cannot identify how the theoretical mechanisms function in practice. To shed light on how the predictions of the model function, I conducted interviews with top trade experts with the European Commission (EC), the USTR, and international law firms. The interviews are not designed to test the models, but instead as a plausibility probe and to understand how private firms engage governments and what effect they have on WTO case selection. All interviewees had engaged in numerous WTO trade disputes and were still actively engaged...
or anticipated being engaged in WTO disputes. The interviewees included a former USTR General Counsel, USTR Legal Advisor in the mission to the WTO, International Trade Counselor to the European Commission, and private international trade attorneys.

Reflecting their varied backgrounds and nationalities, the interviewees highlighted notable variation in how private firms and governments work together, but regardless of whether they were discussing WTO complaints raised by countries in Latin America, Europe, the Asia-Pacific region, or North America, each commented on the important role of private firms in the dispute escalation process. The USTR General Counsel noted that having firms actively engage in the fact-finding and development of the basic legal arguments was invaluable \[\text{(USTR General Counsel 2009)}\]. The other USTR official commented that the agency was “very needy” when it came to preparing the facts of WTO cases and that firms commonly did much of the leg work of the fact finding \[\text{(USTR Official 2009)}\]. Similarly the EC official noted that the EC is ill equipped to independently evaluate and pursue fact intensive cases \[\text{(European Commission Official 2009)}\].

Although the consensus amongst those interviewed is that private firms play a critical role in the dispute settlement process, it was also noted that different cases and countries yield different styles of government-firm interactions. For example, according to a partner at a major international law firm involved with a WTO case involving Brazil (Embraer) and Canada (Bombardier), the government contributed a mere five percent of the total costs, while the private companies paid the remaining 95 percent \[\text{(Trade Attorney 2009)}\]. This example is on the high-end of private party contributions, but the same partner estimated that the average cost breakdown across WTO disputes would be distributed 20 percent to the government and 80 percent paid by private parties. These figures illustrate the reliance of governments on private parties to identify strong cases, develop them, and pursue WTO complaints. The breakdown of costs illustrates that private parties not only play a significant role in identifying harmful violations, but also play a vital role in signaling the strength of
the case and offsetting costs by contributing the majority of litigation expenses in many cases.

Interviews with government officials also provided support to the argument that firms act as a form of fire alarm and are the first movers of the dispute settlement process. Even though the WTO formally blocks private dispute initiation, both USTR officials noted that private parties typically initiate the naming and blaming phases that lead to WTO complaints (USTR General Counsel 2009; USTR Official 2009). Specifically, the former USTR General Counsel noted that private parties are generally responsible for identifying a trade violation, gathering the basic facts of the case, and preparing initial legal arguments, which are then presented to the government to be formally brought to the WTO (USTR General Counsel 2009). He also noted that in the US the agency does not actively seek out potential complaints to pursue in the WTO. Rather than acting as a police patrol the government responds to private companies who “pitch” cases to the government. Prior to the initial pitches the private parties have already substantially contributed to the litigation effort by identifying strong cases and conducting the fact finding and preliminary legal analysis of the case. In this manner, even though firms lack formal access to the litigation process such as in transnational arbitration, they still play a consequential role in the monitoring, enforcement, and litigation of WTO obligations.

The interviews illuminate some of the mechanisms of WTO case selection and lend general support to the theory and broader statistical results. In aggregate, the evidence points to a prominent informational role for private firms in helping governments screen potential cases and enforce WTO obligations. The findings illustrate that industries with dominant firms react to highly distorting trade barriers by contributing to the litigation costs of the case, which leads governments to update their beliefs about the strength of the case and their assessment whether or not to initiate a WTO dispute.
1.6 Conclusion

The theory presented in this paper has direct implications for our understanding of firms’ roles in influencing trade policy. The model shows that enforcement of international trade obligations is significantly influenced by private firms’ role in monitoring and enforcing WTO obligations and suggests that unitary actor models that focus on the formal rules of the organization underreport the number of claims that are profitable for states to initiate. Specifically, cases where costs are only slightly higher than expected profits would be deemed unprofitable under previous models. In contrast, the model presented here predicts that these cases become the most likely cases for firm participation. It is worth noting that the model also predicts that firms will contribute in cases where the gap between costs and profits is larger, provided that the industry has the available resources to contribute and profits still outweigh the litigation costs of the firm. This situation is exemplified by the role of Airbus and Boeing in preparing much of the arguments in the commercial aircraft disputes, where potential payoffs could be as high as a combined $123 billion - the alleged subsidies in the cases (Greene 2007).

The model also suggests that the branches of literature that focus on compliance with international trade law and increasing access to the dispute settlement process for developing countries have overlooked one of the most important mechanisms to achieve their goals. Informal private firm contributions can enhance WTO participation by helping governments effectively screen potential disputes and more efficiently enforce WTO obligations. For scholars interested in increasing developing country participation in the WTO, this paper highlights a mechanism beyond international legal service centers and private sector pro bono legal assistance (Bown and Hoekman 2005). However, even though private firm participation allows states to more efficiently monitor and enforce WTO obligations, without facing the risks of formal access to private dispute initiation associated with transnational dispute settlement, it may also raise new concerns about redistributive consequences. For
example, industries with dominant firms are more likely to overcome collective action challenges, making them more likely to have their interests represented at the WTO, whereas more diffuse industries may find it harder to have their voices represented at the WTO.\footnote{22}

The theory presented here demonstrates the importance of understanding the role of firms for WTO participation and the enforcement of international trade law. While domestic interest groups are often blamed for trade protection, it is clear that private firms also promote trade liberalization by monitoring and enforcing international agreements. In a broader context, this paper contributes to the debate on the monitoring and enforcement of international law and the significance of formal and informal rules and procedures in international organizations. As presented, private firms are significant actors who play a valuable role in shaping how the international legal system functions. Even when formally denied access to dispute initiation, private parties actively engage in the international legal system and play a defining role in how states respond to violations of international law. Although non-state actors were omitted from the original text of the Dispute Settlement Understanding and were left out of much of the early literature on WTO dispute settlement, private parties’ contributions to the litigation process allow governments to screen potential cases and represent a salient path that non-state actors use to effectively engage in and influence the international legal system.

\footnote{22}{I thank an anonymous reviewer for emphasising this point.}
Chapter 2

When to Compromise? The Influence of Compromise, Proposal Power, and Partisanship on Domestic Audience Costs

2.1 Introduction

Hans Morgenthau (1948, 433) famously wrote that “No man who has taken such a stand before the attentive eyes and ears of the world can in full public view agree to a compromise without looking like a fool and a knave.” Yet even as he noted the risks of public compromise in the age of “new” diplomacy, Morgenthau recognized compromise as essential to the maintenance of peace and order in international relations. Almost seventy years after the publication of Politics Among Nations, scholars continue to fear the effects of public compromises. Indeed scholars have written extensively on the costs of inconsistent rhetoric.

See the “Four Prerequisites of Compromise” outlined in Morgenthau’s rules of diplomacy (Morgenthau 1948, 441).
and effects of changing policies on public opinion (Fearon, 1994; Hummel, 2010; Kertzer and Brutger, 2014; Levy et al., 2015; Poole, 2007; Tomz and Houweling, 2012). Although research shows that leaders are generally punished for acting inconsistently, existing work has avoided examining how public opinion responds to compromises, as opposed to general inconsistency. Since theories of political positioning (Hummel, 2010), diplomacy (Morgenthau, 1948; Yarhi-Milo, 2013), and international bargaining (Fearon, 1994; Guisinger and Smith, 2002; Tarar and Leventouglu, 2013) rely on assumptions about how domestic public opinion responds to leaders’ strategies, it is time that we directly examine how public opinion responds to compromise.

The idea that compromise and inconsistency generate domestic political costs is most prominently integrated into IR theory by Fearon’s (1994) seminal work on domestic audience costs. Since Fearon wrote that backing down from threats generates domestic costs due to concerns about “credibility, face, and honor” (Fearon, 1994, 581), domestic audience costs have become common place in crisis bargaining models (Haynes, 2012; Kurizaki, 2007; Tarar and Leventouglu, 2009). Theories and models that incorporate audience costs implicitly rely on psychological microfoundations, generally arguing that audiences punish leaders for saying one thing and doing another due to individuals’ concerns about reputation, competence, or national honor (Clare, 2007; Gelpi and Grieco, 2015; Guisinger and Smith, 2002; Smith, 1998; Tomz, 2007). Yet even as audience costs have received significant attention from both supporters (Davies and Johns, 2013; Schultz, 1998; Tomz, 2007; Trager and Vavrek, 2011; Weeks, 2008) and skeptics (Downs and Sechser, 2012; Snyder and Borghard, 2011; Trachtenburg, 2012), the psychological foundations for the existence of audience costs have received much less attention. Recent experimental studies have shown public opinion shifts in ways consistent with audience cost theory (Davies and Johns, 2013; Levendusky and Horowitz, 2012; Levy et al., 2015; Tomz, 2007), but these studies only examined public opinion in the

\footnote{For a notable exception, see Croco and Gartner (2014) who argue politicians are judged on their current position, as opposed to changes in their position.}
extreme cases where a state makes a threat and then backs down and receives nothing. No study has systematically examined whether public opinion responds to compromises in a manner that supports scholars’ extreme aversion to public compromises, as was epitomized by Morgenthau (1948, 432) who argued “No government which wants to stay in power or simply retain the respect of its people can afford to give up publicly part of what it had declared at the outset to be just and necessary, to retreat from a position initially held, to concede at least the partial justice of the other side’s claims.”

Rather than assuming that the public opposes compromises, this paper systematically examines public attitudes toward compromises, focusing on three key contributions. First, building from a robust experimental literature on inconsistency and audience costs, I employ a series of survey experiments, which illustrate that compromise does not systematically generate disapproval among domestic audiences. Using causal mediation analysis, I further probe the foundations of support for compromise, finding that audiences believe leaders preserve their reputations when negotiating compromises, even if the compromise concedes more from a policy perspective. Second, building from psychological and management theories of leadership, I introduce an original theory of “proposal power,” whereby leaders who are perceived as initiating compromises, as opposed to accepting them, garner greater public support. Proposal power allows leaders to frame the perception of negotiated settlements, such that leaders can avoid domestic punishment for backing down from threats. Lastly, I show that perceptions of compromise and proposal power are significantly influenced by partisan politics. Rather than politics stopping at the water’s edge, partisanship plays a critical role in shaping how audiences react to foreign policy compromises and conditions the diplomatic incentives leaders face at home. Taken together, these contributions show that leaders have significant flexibility to negotiate compromise settlements on the international stage and, contrary to the prevailing wisdom, public compromises do not jeopardize the respect of the public or the leader’s reputation.
Moving away from the assumption that audiences uniformly punish compromise is not just a theoretical endeavor, but is important for capturing many real-world bargaining environments. An example of this type of bargaining scenario played out in the Syrian crisis, where President Obama made a public threat to the Syrian leadership in 2012, but later chose to back down from that threat and instead accept a compromise. The president’s threat was made when he announced that the use of chemical weapons would cross a “red line.” He reiterated the threat in a letter to congress and to reporters in 2013, making it a clear and public threat (Kessler 2013). The threat was broadly publicized by major media outlets as a threat to use military force, with the New York Times publishing the headline “Obama Threatens Force Against Syria” (Landler 2012) and the Wall Street Journal writing “President Threatens Military Response Against Any Use of the Banned Arms” (Entous and Lee 2012). Yet even after the administration reported “definitive proof” that the Syrian regime had used chemical weapons (Entous 2013), President Obama was quick to embrace the idea of a compromise that avoided following through on his threat of force (Borger and Roberts 2013). Instead of following through on a military option, the president supported Russia’s proposal to place Syria’s chemical weapons under international control. Compromises such as President Obama’s in the Syrian crisis are not uncommon and thus deserve our attention when evaluating the constraints under which leaders conduct international negotiations.

This paper proceeds in the next section with a discussion of compromise and how it fits with international bargaining and audience cost theory. Section 3 argues for a reconceptualization of what audiences want in international negotiations, presenting a theory of support for compromise, focusing on proposal power and partisanship. Section 4 outlines the design of the experiments and their advantages and limits, while Section 5 discusses the results and key findings. The paper concludes with a brief discussion of the implications for diplomacy.

\footnote{The compromise agreement regarding chemical weapons was inspired by a comment from Secretary of State, John Kerry, which was “not intended to be a diplomatic opening.” However, before Kerry’s return flight to Washington had even landed, Russia had officially proposed a settlement where Syria would relinquish control of its chemical weapons (Gearan, DeYoung, and Englund 2013).}
bargaining theory, and leaders’ ability to use public escalation as a commitment device or credible signal in international crises.

2.2 Diplomacy, Compromise, and Audience Costs

Whether negotiating a trade agreement or contemplating military action, leaders often have an incentive to push for a bigger piece of the pie. This means leaders will claim to have high demands or initiate threats that misrepresent their true reservation price in negotiations in an effort to gain a better outcome. While asking for more may be a sound bargaining strategy in some situations, it can also build expectations among domestic constituents that may make it harder for leaders to back down from their demands or threats and instead accept compromises. This is why scholars of both international relations (Fearon, 1994; Morgenthau, 1948) and American politics (Hummel, 2010) argue that leaders pay a domestic cost for backing away from their stated positions. This should be especially true in the age of modern diplomacy, where the media provides the public greater access to information about their leader’s diplomatic positions (Potter and Baum, 2010; Baum and Potter, 2014; Slantchev, 2006), thus subjecting diplomacy to the “vice of publicity” (Morgenthau, 1948, 431).

The most prominent theory of foreign policy accountability is Fearon’s (1994) theory of domestic audience costs. The theory assumes that domestic publics impose costs on their leaders if they fail to follow through on public threats. Scholarship invoking domestic audience costs perpetuates the longstanding idea that leaders jeopardize their domestic standing and face public ridicule if they back down from their bargaining positions under the watchful eye of domestic publics (Morgenthau, 1948). Recent research has expanded the theory and shown that domestic publics hold their leaders accountable not only for threats, but also public promises on the international stage (Levy et al., 2015). From these theories, the IR literature generally assumes that once leaders publicly take a stand, they will pay a domestic
cost in the form of lower public approval or lost votes, for any form of inconsistency, such as agreeing to a compromise where they fail to follow through on their original threat or promise.

The implications of domestic public opinion costs for international diplomacy and bargaining are two-fold. First, as Morgenthau (1948) pointed out, international diplomacy requires compromise and subjecting international negotiations to the whim of public opinion jeopardizes international peace and security by undermining the potential for compromise. Leaders must have the capacity to give ground in negotiations, and thus leaders need to believe they can make concessions and compromises without fear of their domestic constituents retaliating against them. Second, domestic audience costs can serve a useful role in international bargaining by acting as a credible signal of leaders’ resolve. The primary value of audience costs is that they can “lock in” the leader, at which point it is more costly for her to back down on her threat or promise than it is to follow through, which makes the leader’s statements credible.

While audience costs theory and public opinion of foreign policy play a critical role in shaping negotiating strategies – they have contributed to a broad research agenda on credible signaling (Fearon, 1997; Mansfield, Milner, and Rosendorff, 2000; Schultz, 1998; Uzonyi, Souva, and Golder, 2012; Weeks, 2008) and are often used to explain why democracies are more peaceful than autocracies (Eyerman and Hart, 1996; Fearon, 1994; Schultz, 1999) – scholars continue to assume audiences disapprove of compromise without testing this fundamental assumption. If audiences support compromises, it opens the door to a re-examination of how we think about bargaining incentives and models of negotiations, with the possibility that leaders can pursue a broader range of peaceful settlements. Introducing the possibility of compromise into an analysis of international negotiations is not a trivial adjustment to existing theories, but instead fundamentally changes how we think about public constraints.
in international bargaining and the implications for credible signaling and leaders’ ability to pursue diplomatic settlements.

To examine some of the implications of how support for compromise affects bargaining incentives, let us consider how domestic constraints shift if we reconsider prevalent assumptions about audience preferences and the nature of audience costs. Audience costs theory assumes that once a leader publicly escalates a dispute she will face heightened costs of backing down and eventually become locked in, such that conceding becomes more costly than going to war (Fearon, 1994). The dichotomous decision to back down or go to war is partly an artifact of the classic crisis bargaining model, shown in Figure 2.2 that shows once a threat is made and the opposition chooses not to concede, the leader must choose to either follow through on the threat and attack or back down. Given the limited options of this scenario, scholars have concluded that audience costs imposed at the “Not Engage” node incentivize the leader to choose to go to war instead of backing down, but scholars have overlooked whether leaders have flexibility to choose some alternative between backing down and not engaging versus going to war.

Figure 2.1: Crisis-Bargaining Model
If we relax the constraints on the leader’s strategies and consider the possibility that after the home leader makes a threat and the foreign leader does not concede, one of them can propose a compromise that the other leader may accept, then the home leader is no longer faced with a simple all-or-nothing decision to wage war or back down. If a compromise is reached, the leader is acting inconsistently by failing to follow through on her threat, which means she may still face audience costs. The question then becomes whether audiences actually impose costs on leaders who reach compromises, as is broadly assumed, and if so whether the costs are great enough to lock in the leader or send a credible signal or resolve?

Two points are worth noting with respect to the signaling value of public threats in the shadow of compromise. First, for audience costs to lock in a leader, approval for the leader must be lower when she initiates a threat and then reaches a compromise, as opposed to following through on the threat and attacking. If public opinion does not punish leaders who compromise, then the leader’s hands are not tied and the public threat does not act as a credible commitment. Second, while audience cost theory predominantly invokes a hand-tying logic, leaders who are not locked in may still pay a cost that could be used to send a credible signal, even though the leaders’ hands are not tied. The relevant comparison to identify such costs is whether compromise is more costly to the leader than staying out of the conflict entirely. If a leader pays a cost for making a threat and then reaching a compromise, relative to staying out of the crisis, then the cost functions as a screening mechanism. Only leaders who have a sufficiently high value for the disputed prize will be willing to initiate the threat. Throughout this paper, I examine both the extent to which public escalation, conditional on the possibility of a compromise, generates costs that lock in leaders and also lesser costs that can be used to signal resolve.

---

4 Compromise can be defined as a situation where a state agrees to concessions that are less than they previously threatened not to accept, but are short of backing down entirely.

5 Tarar and Leventough (2013) also find that audience costs can play a noninformative role through bargaining leverage.
2.3 Reconsidering What Audiences Want

To understand why and how audiences react to their leaders’ bargaining strategies, I first revisit the most prominent explanations for why audiences punish leaders who act inconsistently on the international stage and then, building upon insights from psychology, management studies, and theories of partisanship, develop a new foundation from which to understand perceptions of compromise.

Prominent explanations of why audiences punish leaders for inconsistency invoke concerns about the leader’s or the country’s reputation (Kertzer and Brutger, 2014; Fearon, 1994; Guisinger and Smith, 2002; Levy et al., 2015) or the leader’s competence (Gelpi and Grieco, 2015; Smith, 1998). As Tomz notes, audiences “believe that hollow threats and promises undermine the country’s reputation; that empty commitments are dishonorable and embarrassing; or that inconsistency is evidence of incompetence” (Tomz, 2007). While these mechanisms may condition approval for leaders who back down from threats and do nothing, I argue that compromise is viewed in a substantively distinct manner, such that these beliefs are tempered, or completely absent, when leaders choose not to follow through on a threat and instead negotiate a settlement, especially when the leader proposes the settlement.

In contrast to IR’s general aversion to compromise strategies, scholars of psychology and management studies find that concessions and compromise yield positive outcomes in terms of how audiences feel about negotiations and their perception of the negotiator or leader. While IR scholars generally argue that signaling resolve is important in bargaining (McManus, 2014), management scholars have found that not offering concessions “can create the impression that a focal negotiator does not act ‘in good faith,’ which in turn can result in deadlock...” (Huffmeier and Freund, 2014). For example, in their study of bargaining perceptions and behavior, Benton, Kelley, and Liebling (1972) found that subjects who observed a negotiator offer concessions were more likely to believe the negotiator was “fair”
and the outcomes were perceived as more satisfying than when the negotiator stayed firm. Furthermore, they found that when the negotiator made a high demand and stuck to it, respondents believed the negotiator was “unfair” and “bad,” although they were also more likely to say the negotiator was “strong.” These findings are consistent with theories of graduated reciprocation in tension-reduction models \cite{Osgood1962}, and studies have found that backing away from demands also makes agreement more likely in negotiations \cite{Benton1972}. Concessions and compromise thus play a vital role in the resolution of disputes and leaders who embrace these strategies are more likely to be viewed as fair. Rather than jeopardizing the leader’s reputation, backing down from demands and reaching a compromise can generate positive approval from audiences and increased satisfaction with the bargaining outcome.

Hypothesis 1: Even when holding policy outcomes constant, leaders who back down from threats and reach compromises will have greater approval than those who back down and choose not to engage.

**Proposal Power**

Of course, there is a potential trade-off that leaders face when backing down from threats and reaching compromises. Leaders who make concessions are less likely to be viewed as strong, even though they are more likely to be viewed as fair and good leaders \cite{Benton1972}. Given that international relations is argued to be dictated by power, or perceptions of power, leaders ought to take steps to preserve their reputations for strength. I argue that leaders can exercise proposal power to enhance audience perceptions of their leadership strength and bolster support for negotiated compromises.

Proposal power – the framing of a negotiated settlement such that the leader is viewed as initiating the agreement, as opposed to accepting it – shapes audiences’ perceptions of leaders’ handling of negotiations. According to leadership theories, leadership requires initiative, coordination, and direction \cite{Bass1990}. When leaders take initiative, audiences and
followers tend to give them higher ratings (Van Vugt 2006). Experimental studies of leadership have found that those who take initiative are more likely to be nominated as leaders in the future (Kremer and Mack 2013), suggesting that voters are more likely to continue to support and re-elect politicians who take the initiative and propose agreements. These theories of leadership suggest that when audiences view politicians taking the initiative, as is the case when they exercise proposal power, the leader fulfills the audience’s expectation for what a leader ought to do, which builds and reinforces the perception of the politician as an effective and strong leader. Consistent with theories of leadership, proposal power allows politicians to preserve their reputations and maintain their support among domestic audiences, even when choosing not to follow through on their threats and promises.

Hypothesis 2: Leaders who propose compromises will have greater approval than those who accept foreign leaders’ compromise proposals.

**Partisanship**

While psychology, leadership, and management studies give us reason to believe that compromise strategies can generate broad support from the public, in the political realm we should also expect partisanship to play an important role in shaping how audiences respond to compromise. Partisanship plays two distinct functions in shaping public opinion of foreign policy. First, partisanship is a salient reference point for domestic audiences that allow them to easily seek out cues from like-minded elites, which shapes how audiences view political events (Berinsky 2007; Cohen 2003; Nicholson 2012). Second, recent scholarship on partisanship suggests that partisan identification reflects deeper individual traits that are likely to affect how people view the world and react to foreign policy (Jost et al. 2003a, 2007; Jost, Nosek, and Gosling 2008). Taken together, partisanship plays an important role in shaping how audiences respond to leaders who engage in compromise.

Partisanship plays a key function in shaping foreign policy opinion by providing salient groups from which audiences take cues and filter information. Given that the general public
is often poorly informed when it comes to foreign policy (Holsti, 1996), audiences are likely to turn to their political leaders and respected elites for cues that shape their foreign policy opinions (Berinsky, 2007; Saunders, 2015). In the context of international negotiations, audiences will look for in-group elites from whom to take cues – the most salient elite cues may come from being in the national in-group or their political party in-group. When evaluating negotiated settlements, audiences will focus on whether a compromise is proposed by the leader of their country, since international negotiations inevitably involve an “us” versus “them” divide, where audiences are more likely to accept and support compromises proposed by their country’s leader. The audience will also look to salient political reference groups, generally defined by party-affiliation. If the leader who initiates the proposal is a member of their party, then the compromise is proposed by a leader who is both a member of their party and their country, making them the most likely audience to support the compromise settlement. Given these overlapping reference groups for cues, audiences are more likely to support compromises when their country’s leader proposes the settlement, and even more so when the proposer is a member of their own party.

Hypothesis 3: Audiences will have greater approval for leaders of their own political party who reach compromises, as opposed to those from opposing parties.

Partisanship identification also acts as a salient sorting mechanism in domestic politics. In the United States, liberals are more likely to identify as Democrats and conservatives are more likely to be Republicans and similar sorting occurs in other countries. More importantly, the liberal-conservative spectrum is highly correlated with individuals’ world-views that are salient when determining how they form beliefs about foreign policy. A growing literature on the psychological foundations of political ideology finds that across countries conservatives are more resistant to change, oppose ambiguity, and prefer definitive outcomes that provide closure (Jost et al., 2003a, 2007; Jost, Federico, and Napier, 2009; Zavala, Cislak, and Wesolowska, 2010). This suggests that conservatives, who are more likely to
resist change and oppose ambiguity, should be the most likely to disapprove of leaders who change their minds or shift strategies. Additionally, conservatives are likely to view the world as more threatening, and are thus more likely to prefer aggressive responses, such as military action (Jost, Federico, and Napier 2009; Zavala, Cislak, and Wesolowska 2010). Each of these traits should lead Republicans to be more skeptical of compromises in crises than Democrats. In contrast, liberals are more likely to be accepting of uncertainty (Jost, Federico, and Napier 2009), and to be accommodationist and support cooperative internationalism (Holsti and Rosenau 1993, 1996). Sorting of liberals and conservatives based on these traits occurs in the U.S., where Democrats are found to be more dovish than Republicans (Herrmann, Tetlock, and Visser 1999; Petrocik 1980; Zavala, Cislak, and Wesolowska 2010), which suggests that Democrats should have more favorable views of compromise than Republicans. These individual traits and perceptions of the world provide a framework for understanding how partisanship shapes public approval of compromise. Taken together, these traits suggest that Democrats should be more likely than Republicans to support leaders who shift strategies away from threats and instead reach compromises.

Hypothesis 4: Democrats will have higher approval for compromise than Republicans.

2.4 Experimental Methods

To analyze the micro-foundations of how audiences respond to leaders’ strategies in international crises, I build on previous experiments that use the classic “repel an invader” scenario. Experimental studies are particularly well suited to isolate changes in public approval in crisis scenarios, given that they allow the researcher to hold the crisis context constant across a range of strategies and outcomes, ensuring internal validity of the study. Recent experiments have found audience costs to be present across a range of contexts and countries, providing a
robust literature on which I build (Davies and Johns, 2013; Levendusky and Horowitz, 2012; Tomz, 2007; Trager and Vavrek, 2011). Tomz employed the first survey experiment to test the foundation of audience costs theory and found that audience costs exist across a range of leaders’ threats (Tomz, 2007). Trager and Vavrek (2011) expanded upon Tomz’s analysis and tested the scope conditions and micro-foundations for the audience cost argument across a broader range of crisis scenarios. However, both experiments provided a narrow analysis of bargaining strategies, where the president’s strategy was always an extreme. While a useful approach for evaluating whether audience costs exist in the most likely (extreme) cases, each experiment leaves out a broad range of relevant strategies. No experiment has examined how audiences react when a leader escalates a crisis, fails to follow through on a threat, and then negotiates a settlement or compromise. I show that even where previous studies found significant audience costs, those costs are mitigated or eliminated when compromise strategies are evaluated.

An additional advantage of using survey experiments to test the effects of compromise and proposal power, is that survey experiments represent the most likely case to observe audience costs, making them a harder case for the ameliorating effects of compromise and proposal power. In traditional audience cost experiments the leader’s threat and subsequent inconsistency are temporally proximate and thus audiences should be most likely to punish leaders who back down, whereas in real world settings audiences may have more difficulty connecting a past threat to the leader’s current choice of action. Furthermore, respondents are not presented with competing messages about the efficacy or value of the engagement, which means respondents are more likely to focus on the leader’s threat and subsequent

---

6 Trager and Vavrek (2011, 527) also tested the effect of partisanship on audience costs and found no partisan effects; however they did show that partisan cues played an important role when leaders chose to go to war.

7 Trager and Vavrek (2011, 528) are aware of this omission and suggest it is an area deserving of future research.

8 See Levendusky and Horowitz (2012) for an exception, where respondents are presented with new information justifying the change in policy.
inconsistency. For each of these reasons, audience cost experiments represent the most likely situation for leaders to be punished, making it an ideal manner in which to test whether compromise and proposal power can overcome the negative effect of leaders’ inconsistency.

Using a similar framework as previous studies, I fielded a series of online survey experiments. Experiment 1 was fielded in the spring of 2014 on a sample of 916 respondents. Experiment 2 was fielded in spring of 2013 on a sample of 1761 respondents. Both studies recruited respondents using Amazon’s Mechanical Turk (mTurk). Samples from mTurk are increasingly widely used in both political science and elsewhere in the social sciences because of the extent to which they can replicate classic experiments on more diverse samples than those traditionally employed in political psychology research, as recent review pieces by Berinsky et al. (2012), Buhrmester, Kwang, and Gosling (2011), Horton, Rand, and Zeckhauser (2011), Paolacci, Chandler, and Iperiotis (2010), and Rand (2012) make clear.

As is common in online surveys, my samples are more male, educated and liberal than the American population as a whole – 56.2% of respondents were male, 58.9% have at least a four-year college degree, and 51.3% identify themselves as being liberals in Experiment 2. The samples are particularly useful, since I focus my analysis on specific partisan subgroups, and thus draw conclusions contingent on the relevant subgroup. Additionally, given that the mTurk samples successfully replicate the audience cost effect found in previous experiments across studies in the United States and abroad, they generate the necessary baseline to examine changes in audience costs. While Berinsky et al. (2012) and others have shown the reliability of mTurk results, I took the additional step of fielding a truncated version of the experiment on a national sample administered by Qualtrics in the fall of 2012. The three experiments find consistent results across three years of study and different sample populations, showing the results are not a function of time or sample method.

---

9 Demographics for all samples are included in the appendix, section B-1.
10 Additional details regarding the national sample and a comparison of core results across studies are provided in the appendix, section A-2. The Qualtrics study is not used as the primary analysis, because the experiment does not control for policy outcomes such as casualties and territorial outcomes.
The experiments began with a script that read:

“The following questions are about U.S. relations with other countries around the world. You will read about a situation our country has faced many times in the past and will probably face again. Different leaders have handled the situation in different ways. We will describe one approach U.S. leaders have taken, and ask whether you approve or disapprove.”

In both experiments, respondents next read about a crisis where a country sent its military to take over a territorial region in a neighboring country. Consistent with previous experiments on audience costs, the scenarios deliberately avoided using any country names for the foreign parties. This was done to avoid country-specific confounders and to maintain comparability with previous studies. After reading the background on the crisis, respondents were randomly assigned the president’s strategy. In the baseline condition, labeled Stay Out, the president does not engage in the crisis.

- Stay Out: “The U.S. president, who was a Democrat (or Republican), said the United States would stay out of the conflict. The attacking country continued to invade and the conflict ended with the attacking country taking control of 20 percent of the contested territory.”

In each of the remaining treatment conditions the U.S. escalated with a threat and then the respondent was randomly assigned to the Compromise, Engage, or Not Engage condition:

Threat: “The U.S. president, who was a Democrat (or Republican), said that if the attacking country continued to invade, the United States military would immediately engage and attempt to push out the attacking country. The president sent troops to the region.”
• Compromise Condition: “The attacking country continued to invade, but the president did not immediately engage. The U.S. president (or leader of the attacking country) proposed a settlement, which was agreed to by all parties, and the conflict ended with the attacking country taking control of 20 percent of the contested territory.”

• Engage Condition: “The attacking country continued to invade and the U.S. president ordered the military to engage. The U.S. did not lose any troops in the conflict and the conflict ended with the attacking country taking control of 20 percent of the contested territory.”

• Not Engage Condition: “The attacking country continued to invade. The U.S. president ordered the military not to engage. The attacking country continued to invade and the conflict ended with the attacking country taking control of 20 percent of the contested territory.”

As in similar studies, each respondent is presented with only one of the potential conditions. For those respondents in the Compromise condition, there is an additional nested treatment that specifies who proposed the negotiated settlement. After reading the president’s choice of action, respondents were provided a summary of the events. Respondents were then asked whether they “approve, disapprove, or neither approve nor disapprove of the way the U.S. president handled the situation” and then how strongly they felt that way. These responses generated the dependent variable, which is an approval score on a seven point scale ranging from -3 being strongly disapprove to +3 being strongly approve.

A challenge of evaluating public approval for leaders’ strategies in real-world crises, and even in some experiments, is that audiences are likely to have preferences over both the

\[11\]In the truncated Qualtrics version of the experiment, the Compromise treatment wording did not include “which was agreed to by all parties,” but the results remained consistent. I also tested audience reactions when the conflict ended with the attacking country taking control of 50 percent of the contested territory. The results were not significantly different when the compromise gave the attacking country 50 percent of the territory – a results that is discussed in greater detail in the results section.
strategy used by the leader and the policy outcomes of the crisis (Chaudoin 2014b; Snyder and Borghard 2011). While disaggregating the audience’s preference for outcomes and strategies is challenging in observational studies, using an experimental scenario allows the magnitude of the audience cost to be measured while holding policy outcomes constant. In order to isolate the audience cost across all strategies, as noted in the treatments above, I control for the outcomes across all conditions in Experiment 1. This allows me to first test how the president’s strategy alone affects the magnitude of audience costs. I then analyze nested treatments in Experiment 2 to test whether approval for compromise is conditioned by proposal power and whether additional concessions that lead to worse policy outcomes influence approval.

2.5 Results

The average approval scores, conditional on the party of the president, for both Experiments are reported in Figure 2.2. These results hold the outcome of the crisis constant, with the foreign country taking over 20 percent of the territory in each scenario, isolating the effect of the president’s strategy. The first notable result is that both experiments replicate previous studies’ findings that audience costs exists for leaders in the Not Engage condition, where the leader makes a threat, does not follow through on it, and has no further involvement in the crisis. The Not Engage condition consistently receives the lowest approval score regardless of whether the president is a Republican or Democrat. These findings suggests that the experiments capture the baseline audience cost effect as in previous studies (Davies and Johns 2013; Levendusky and Horowitz 2012; Tomz 2007; Trager and Vavrek 2011), which

12 Casualties are held constant at zero. A further discussion of the importance of holding casualties constant is provided in the appendix, section B-3.
13 The average approval scores not conditioned on party of the president are shown in the appendix, section B-4.
is a necessary starting point given that the following results focus on mitigating audience costs through compromise.

**Figure 2.2: Average Approval Score for Each Strategy**

*With a Republican as President*

![Graph showing average approval score with Republican President.]

**Figure 2.2** shows the average approval score and 90 percent confidence intervals for experiments 1 and 2 for each randomly assigned crisis strategy and party of the president. Outcomes are held constant with no casualties and 20% territory gained by the attacking country. The approval score is measured on a scale from -3=Strongly Disapprove to 3=Strongly Approve.
2.5.1 The Effect of Compromise

Figure 2.2 also shows that regardless of the president’s partisanship, and even when holding outcomes constant, audiences prefer leaders who back down from their threat and negotiate a compromise, as opposed to those who simply choose not to engage after having made a threat, providing support for hypothesis 1. While these general results are a useful starting point, they obscure the effect of partisan audiences since they do not account for the interaction of the audience and the President’s partisanship. To fully examine partisanship and how different audiences react to compromise, I begin by comparing approval for the Compromise condition to the Not Engage condition. Table 2.1 presents the results, conditional on both the party of the audience and the president. For each combination of partisan audience and party affiliation of the president, average public approval is higher in the Compromise condition than the Not Engage condition. The difference in means between these conditions is always positive and is statistically significant for audiences of Democrats, while the results are positive but not significant for Republicans. Without taking into consideration the president’s proposal power, which will be discussed in subsequent sections, Table 2.1 suggests that audiences, particularly Democrats, prefer leaders who reach compromise settlements as opposed to not engaging after a threat, which supports both hypotheses 1 and 4.

To examine why audiences tend to favor compromise, I focus on one of the most prominent explanations for why audiences choose to punish leaders’ inconsistency – reputational concerns (Kertzer and Brutger, 2014; Fearon, 1994; Guisinger and Smith, 2002; Weeks, 2008; Tomz, 2007). To measure concerns about reputation, each respondent was asked “On a scale of 1-5, how much damage do you think there would be to the President’s reputation as a result of the President’s handling of the situation?” When leaders chose to reach a compromise,

---

14 The full results for Experiment 1 and Experiment 2 are reported in the appendix, section B-5.
15 Results for Independents are reported in the appendix, section B-6.
Table 2.1: Experiment 1 - Differences in Approval Between Not Engage and Compromise

<table>
<thead>
<tr>
<th>Party of Audience</th>
<th>Republican</th>
<th>Democrat</th>
</tr>
</thead>
<tbody>
<tr>
<td>President is Democrat</td>
<td>0.39</td>
<td>1.57***</td>
</tr>
<tr>
<td></td>
<td>(0.38)</td>
<td>(3.94)</td>
</tr>
<tr>
<td>President is Republican</td>
<td>1.07</td>
<td>1.34***</td>
</tr>
<tr>
<td></td>
<td>(1.37)</td>
<td>(3.15)</td>
</tr>
</tbody>
</table>

Statistical significance: *p<0.1; **p<0.05; ***p<0.01

Table 2.1 shows the difference in average approval scores (Compromise – Not Engage) and t-statistics among Democrats and Republicans based on the party of the president while holding territorial and casualty outcomes constant at no casualties and 20% territory gained by the attacking country. The approval score is measured on a scale from -3=Strongly Disapprove to 3=Strongly Approve.

after backing down from a threat, as opposed to not engaging, they received a significant improvement in their perceived reputation (0.65, p = 0.00). These results are confirmed using nonparametric mediation analysis (Imai et al., 2010) to test how perceptions of reputation affect audiences’ approval of the president’s handling of the situation. When examining the increase in approval from the Not Engage to the Compromise condition, the mediating effect of reputation is responsible for 61 percent of the increase in approval (p = 0.00). These findings demonstrate that leaders who are concerned about their reputations among domestic audiences, can preserve or enhance their reputations by reaching diplomatic compromises.

2.5.2 Sensitivity to Outcomes

While audiences favor leaders who negotiate compromise settlements over leaders who back down and don’t engage, the next question is whether these results are robust to variations in crisis outcomes. Snyder and Borghard argue that the substantive policy outcomes of a crisis play a significant role in determining public approval, which could undermine the mitigating effect of compromise on audience costs if leaders are punished for making additional con-

\[16\] The mediation results are shown in the appendix, section B-7.
cessions (Snyder and Borghard, 2011). To evaluate this concern, in Experiment 2, I tested the sensitivity to of audiences’ approval to randomly varied crisis outcomes. Experiment 2 follows the same structure as the first experiment, but the amount of territory gained by the attacking country was varied. I also randomly assigned the number of casualties within the Engage condition, with the expectation that additional casualties will decrease approval for engagement (Gartner, 2008; Gartner and Segura, 1998; Gelpi, Feaver, and Reifler, 2005). These nested treatments allow me to examine how sensitive audiences are to different types of crisis outcomes, which also helps to further illustrate the differences between Democrat and Republican audiences.

I find that audiences do not punish leaders for worse outcomes in the Compromise condition. The outcomes of the negotiated settlement led to the foreign country gaining a baseline of 20 percent of the contested territory or an additional 30 percent, which emphasizes the leader’s concessions when compromising. Among Democrats the shift in average approval when the attacking foreign country gained an additional 30 percent of the territory was essentially zero at 0.11 ($p=0.72$) and among Republicans it was -0.45 ($p=0.34$). One explanation for the lack of sensitivity to outcomes in the Compromise condition is that respondents don’t have a strong preference for outcomes unless the United States is militarily engaged, which raises the salience of the issue and invokes the national honor. This suggests that when the president does not engage militarily she can exercise significant freedom in negotiating outcomes given that the public does not appear to distinguish between varying territorial outcomes.

In contrast to the Compromise condition, audiences of Democrats are sensitive to outcomes in the Engage condition. Table 2.2 shows the effects of changes in territorial outcomes and casualties on public approval in the Engage condition. The outcomes of the military engagement led to the foreign country gaining either none or 20 percent of the contested ter-

\footnote{A discussion of Experiment 2 and the full treatment conditions are provided in the appendix, section B-8.}
ritory and the U.S. experienced either no casualties or 1,000 casualties in the engagement. In
the event the foreign country did not gain any territory, Democrats’ approval score dropped
by 1.03 ($p=0.01$) when there were 1,000 casualties in comparison to no casualties. The effect
of the foreign country gaining 20 percent of the territory (with no casualties) is similar, with
a decline in approval score among Democrats of 1.13 ($p<0.01$). Republicans on the other
hand did not show a significant change in approval in response to crisis outcomes and in-
stead consistently support military engagement. These results support the characterization
of Republicans as consistently hawkish, whereas Democrats’ approval of military strategies
is highly dependent on how favorable an outcome is achieved.

Table 2.2: Experiment 2 - Differences in Approval Based on Outcomes

<table>
<thead>
<tr>
<th>Party of Audience</th>
<th>Republican</th>
<th>Democrat</th>
</tr>
</thead>
<tbody>
<tr>
<td>Casualties</td>
<td>-0.35</td>
<td>-1.03***</td>
</tr>
<tr>
<td></td>
<td>(-0.51)</td>
<td>(-2.63)</td>
</tr>
<tr>
<td>Territory</td>
<td>-0.35</td>
<td>-1.13***</td>
</tr>
<tr>
<td></td>
<td>(-0.51)</td>
<td>(-3.02)</td>
</tr>
<tr>
<td>Casualties and Territory</td>
<td>0.40</td>
<td>-1.84***</td>
</tr>
<tr>
<td></td>
<td>(0.77)</td>
<td>(-4.74)</td>
</tr>
</tbody>
</table>

Statistical significance: *$p<0.1$; **$p<0.05$; ***$p<0.01$

Table 2.2 shows the difference in average approval scores for the Engage condition based on the crisis outcomes, showing the effects of 1,000 casualties, the hostile country gaining 30% more territory, or both, along with the t-statistics among Democrats and Republicans. The approval score is measured on a scale from -3=Strongly Disapprove to 3=Strongly Approve.

2.5.3 Proposal Power

A final component of compromises that is likely to impact audiences’ approval is the percep-
tion of the leader’s role in reaching a settlement, specifically whether they exercise proposal
power. To test the effect of proposal power, Experiment 2 randomly assigned whether a
compromise was proposed by the president or the foreign leader.
The power of being the proposer generally plays in the favor of the president, but partisanship plays a significant role in conditioning the audience’s response. To measure the role of partisanship, I analyze proposal power across different combinations of the president’s party and the audience’s party. I use a “Party Match” variable that indicates whether the audience member and president are from the same party. As is shown in Figure 2.3 Democrats consistently have favorable opinions of presidents who are Democrats and agree to compromises, regardless of whether the president proposes or accepts the compromise. The only situation where Democrats look unfavorably on compromise is when the president is a Republican and the compromise is initiated by the foreign leader. In contrast, and consistent with hypothesis 4, Republican audiences are more selective when supporting compromise. Republicans strongly support a Republican president when she proposes the compromise, but not when she accepts a compromise initiated by the foreign country. Consistent with hypotheses 2 and 3, these findings support the idea that audiences take cues from their leaders in international negotiations, resulting in greater support for the leader when she is a member of the their party and when the leader initiates the compromise. The results also confirm a long-held idea that “only Nixon could go to China,” meaning that Republican presidents, who are more likely to be viewed as hawks, actually have an easier time making peace (Cowen and Sutter, 1998; Schultz, 2005; Trager and Vavrek, 2011). A Republican president who proposes a compromise is strongly supported by both Democrats and Republicans, unlike a Democratic president who proposes a compromise, who only receives strong support from her own party.

2.5.4 Signaling and Lock-In

Given the interaction between proposal power and partisanship, I now look closer at when leaders are locked in by audience costs and when lesser costs are generated that can still be used to send credible signals. I first examine whether audiences of Democrats lock in
Figure 2.3: Average Approval Score for Compromise Based on Party-Match and Proposal Power

Figure 2.3 shows the average approval score and 90 percent confidence intervals for audiences of Democrats and Republicans based on whether the respondent and president are from the same party (party-match) interacted with whether the U.S. president or the foreign leader proposed the negotiated settlement that was then agreed to by all countries, while holding crisis outcomes constant. The approval score is measured on a scale from -3=Strongly Disapprove to 3=Strongly Approve.

Democratic presidents by comparing approval scores in the Engage and Compromise conditions when the president is the proposer. The results show no difference in the approval scores (0.08, $p=0.87$), which means compromise is not more costly than following through on a threat of war and thus the leader’s hands are not tied. Not only does lock-in not occur, I find that there is not a significant difference in approval between Stay Out and Compromise when a Democratic president proposes (-0.44, $p=0.19$), demonstrating that public escalation does not generate costs among Democratic leader’s primary constituents and thus public escalation does not lead to lock-in or credible signals of resolve for Democratic leaders concerned with their party-constituents.

Although Republicans are more wary of compromise than Democrats, Republican audiences still support their president when she proposes a compromise even after backing down
from a threat. In fact, Republicans strongly support Republican presidents in such situations and their average approval is as high as if the president of their party had chose to stay out of the crisis entirely. These findings show that when a Republican president proposes a negotiated settlement after not following through on a threat, the president does not pay a cost of any kind and is not locked in or able to use public approval to send costly signals. This suggests that Republican presidents can strategically back down from threats and negotiate settlements without the fear of punishment from their constituents, but unlike Democrats, they will be punished by their constituents if they agree to a compromise proposed by the foreign leader.

In cases where the president may be concerned not just with support from her party, but with support from both parties, then backing down from a threat and proposing a compromise can be costly, but only for a Democratic president. While a Democratic president who proposes a compromise receives support from her party, she pays a cost among Republican audience members. Depending on the electoral and political environment, this cost may be enough for the leader to send a credible signal of resolve. In contrast, Republican presidents receive strong support from both parties when they propose compromise, and are thus unable to generate costly signals in the shadow of compromise.

2.6 Conclusion

The evidence put forth in this paper builds a case for a reformulation of how we think about public opinion and threats in international bargaining. Three critical points emerge for understanding the role of domestic audiences in international bargaining. First, is that leaders can influence the level of public approval when negotiating a compromise by exercising proposal power. When leaders negotiate settlements and are perceived as the proposer they are able to protect their reputations, which allows them to mitigate or eliminate public disapproval. This introduces a new component to the bargaining dynamic that allows leaders
to frame the bargaining outcome for their home audience and provides greater flexibility to
the leader to negotiate concessions. Further research is needed to evaluate whether leaders’
proposal power holds in other negotiating environments and across issues areas, but this
paper puts forth a strong case for considering how settlements are framed and the role of
proposal power in international relations.

The second takeaway is that the composition of a leader’s audience is of vital importance
when evaluating whether publics constrain leaders behavior and can create costly signals
on the international stage. Whereas previous studies found that audience costs were non-
partisan (Levendusky and Horowitz, 2012; Trager and Vavrek, 2011), I find that partisanship
significantly affects approval for compromise when both the audience and the leader’s par-
tisanship are taken into account. Republican presidents have greater leeway to negotiate
compromise settlements and receive strong support from audiences of Democrats and Re-
publicans when they propose compromises. In contrast, presidents who are Democrats are
more constrained and only receive support from their own party when they propose settle-
ments. The role of partisanship in shaping public approval for crisis strategies adds further
support to a growing literature which argues that increasingly politics does not stop at the
water’s edge (Ahn, 2011; Cavari, 2013; Trubowitz and Mellow, 2005).

Lastly, when we include the possibility of compromises, audiences rarely lock in lead-
ers in international crises and audience costs are, at best, inconsistent signals of leaders’
resolve. Rather than assuming publics disapprove of compromise, as Morgenthau (1948)
argues, these results fundamentally change how we think domestic audiences respond to
international diplomacy. In the Syrian case, audience costs theory would predict that the
president’s public threat would generate domestic audience costs and that the decision not to
follow through on the threat would result in the president being punished. While some elites
criticized the president and argued that the failure to follow through with military action
hurt the president’s credibility (Sink, 2014), the general public was supportive of the com-
promise. Public approval for the president’s handling of foreign policy showed no decline after the president failed to follow through on his threat. According to CBS and NBC/Wall Street Journal Polls, the president’s foreign policy approval ratings were 43 and 46 percent respectively in the months before Syria used chemical weapons and stayed relatively consistent at 45 and 46 percent following Syria’s use of chemical weapons (Polling Report, 2014). When specifically asked about President Obama’s handling of foreign policy toward Syria, 56 percent of Republicans, 80 percent of Democrats, and 65 percent of Independents supported the decisions not to engage in airstrikes and instead pursue a compromise strategy (Pew Research, 2013). If the president’s only options were to back down or attack, then the president would have likely been punished if he chose not to attack, but the compromise option allowed the president to choose not to engage while also minimizing disapproval by the general public. This case illustrates the importance of evaluating how public opinion responds to negotiations that go beyond the traditional crisis bargaining model and shows how the potential for compromise can prevent leaders from tying their hands in international crises.

While President Obama issued a prominent public threat in the Syrian crisis, he was able to accept a compromise settlement and avoid being punished by the general public. Given that the president was generally perceived as accepting a foreign leader’s proposal, this example is a least likely case for compromise to successfully mitigate audience costs, yet audiences across the political spectrum supported the President’s decision not to follow through on his threat. However, partisanship still played a key role in shaping public opinion, evidenced by the 24 percentage point higher approval among Democrats than Republicans for President Obama’s choice to pursue a compromise. This example highlights the role of compromise in allowing leaders to avoid domestic audience costs and the importance of reconsidering how publics react to international negotiations. Rather than constraining

---

18 While elite cues play an important role in shaping public opinion (Berinsky, 2007; Saunders, 2015), in this case the general public did not follow the cues of the more hawkish critics, such as Panetta (Sink, 2014).
leaders’ diplomatic options and jeopardizing international peace and security (Morgenthau, 1948), domestic public opinion often supports leaders who make concessions, which limits the signaling value of public threats and suggests that leaders have significant flexibility to negotiate international compromise settlements.
Chapter 3

Decomposing Audience Costs: Bringing the Audience Back Into Audience Cost Theory

3.1 Introduction

How are leaders able to credibly signal their intentions in foreign policy crises, despite considerable incentives to bluff? According to a considerable volume of research in International Relations, one technique governments have at their disposal is the creation of domestic audience costs: by publicly threatening to use force in order to change an opponent’s behavior, leaders can effectively tie their own hands, since their domestic constituencies punish governments who say one thing and do another (Fearon 1994; Baum 2004; Slantchev 2006; Weeks 2008; Tarar and Leventouglu 2009). In the nearly twenty years since Fearon’s seminal article, a large volume of scholarship has explored the scope and examined the logic of audience costs. Experiments have been central to this endeavor (e.g., Tomz 2007; Hoffman et al. 2009; Trager and Vavrek 2011; Levendusky and Horowitz 2012; Davies and Johns
both because of their ability to avoid the selection effects that make audience costs difficult to study with observational data — if audience costs are real, strategic leaders should avoid making the empty threats that would incur the public’s wrath in the first place (Schultz, 2001; Tomz, 2007) — and because at least in democracies, audience cost models rely on a particular set of assumptions about what publics want (Weeks, 2008).

This body of research has told us at least two things. First, audience costs are about inconsistency: whether because of instrumental concerns about the country’s reputation or normative concerns about national honor, publics dislike leaders who say one thing and do another (Fearon, 1994; Tomz, 2007). Second, audiences are homogenous: the kinds of individual-level characteristics that routinely make a large difference in the study of public opinion about foreign policy seem to play relatively little role in moderating the impacts of treatments in audience cost experiments (Tomz, 2007; Levendusky and Horowitz, 2012; Davies and Johns, 2013). Existing work has thus assumed that audience costs abide by an essentially unitary logic, and that leaders face an essentially unitary audience.

We argue here that both these assumptions are wrong. Because of how audience costs are defined — the decrease in public support for a leader who threatened the use of force and did not follow through, compared to a leader who simply stayed out — audiences can punish the leader for saying one thing and doing another, but also for threatening the use of force in the first place, a sunk cost we refer to here as a belligerence cost. Failing to take the possibility of belligerence costs into account means we risk misreading what these classic experiments are telling us, and misunderstanding the heterogeneous nature of public responses to crisis bargaining. We present a novel experimental design where we disentangle these two competing logics of why audiences punish, and turn to a series of dispositional characteristics from political psychology and the public opinion about foreign policy literature to bring the audience into audience cost theory. Our results suggest three main points. First, there is
no unitary logic of audience costs, as audiences punish leaders both for being inconsistent
and for threatening force, such that traditional audience cost experiments may be overes-
timating how much the public cares about inconsistency. Second, different types of people
demand audience costs for very different reasons: hawks and conservatives, for example,
punish the President for being inconsistent, while doves and liberals punish the President
predominantly for threatening to use force. Both inconsistency costs and belligerence costs
can be broadly thought of as types of audience costs, in that the audience negatively reacts
to the leader’s handling of a foreign policy crisis, but only inconsistency costs follow the
traditional audience cost logic put forth by Fearon (1994), whereas belligerence costs invoke
a different logic altogether. Third, the presence of these two distinct signaling mechanisms
has important implications for crisis bargaining, from explaining the rarity of public threats,
to showing how the credibility of threats depends not just on the content of the President’s
actions, but also on the character of the President’s constituency.

3.2 Audience Costs, Revisited

Audience cost experiments test a proposition derived from classic crisis bargaining games,
whose structure is illustrated in the game tree in Figure 3.1. The game has two players,
represented here as players 1 and 2, who we can think of as the government of a state and
its foreign challenger, although the game can also be used to model threats in negotiations
more generally. Player 1 decides whether to threaten the use of force against player 2;
if player 1 chooses not to threaten force, the game ends with player 1 deciding to stay
out. If, on the other hand, player 1 issues a threat, player 2 must decide whether or not
to concede. If player 2 does not concede, player 1 then must decide whether to follow
through on the threat and engage player 2, or to back down and not engage. Audience
cost models are chiefly concerned with the difference between staying out and not engaging;

\[1\text{See, e.g., Bueno de Mesquita and Lalman 1992, Fearon 1994, Guisinger and Smith 2002}\]
as long as domestic audiences prefer the former to the latter, leaders have to be concerned about domestic political repercussions when they renege on their public commitments. The presence of these costs renders public threats informative rather than merely cheap talk: if leaders expect their domestic constituency will punish them if they issue a threat and back down, they will only issue public threats if they actually intend to follow through; foreign challengers, aware of the existence of audience costs, thus will concede before the crisis can escalate further.

Figure 3.1: Classic Crisis-Bargaining Model

![Diagram of Classic Crisis-Bargaining Model]

Note: Payoffs for each player are presented under each node; the audience cost ($-a$) is the difference in the leader’s approval when she threatens force and backs down (Not Engage), versus not threatening force in the first place (Stay Out).

The validity of these assumptions has come under both empirical and theoretical scrutiny, both from historically-minded scholars failing to find evidence for audience cost logic in real world crises, and experimentalists showing that there are circumstances in which leaders are in fact rewarded for backing down [Brown and Marcum 2011; Snyder and Borghard 2011; Levendusky and Horowitz 2012; Trachtenburg 2012; Mercer 2012]. For our purposes, however, two considerations are especially worth noting here. First, although scholars use multiple mechanisms to explain the existence of audience costs, from concerns about national honor or reputation, to signals of leaders’ quality or competence, embedded in both accounts is the notion that audiences have preferences for consistency: that the public punishes leaders
when their words do not match their deeds (Fearon, 1994; Smith, 1998; Chaudoin, 2014b). Indeed, one of the major charges levied against audience cost theory by its critics is that audience cost models erroneously assume the public cares about consistency rather than the policy outcome itself (Clare, 2007; Snyder and Borghard, 2011; Downs and Sechser, 2012).

Second, although a large body of literature has investigated the scope of audience cost theory by focusing on the nature of the domestic political institutions in which the audience is embedded — for example, whether audience costs vary with regime type, opposition criticism, or the leader’s rhetoric (Slantchev, 2006; Weeks, 2008; Baum, 2004; Levendusky and Horowitz, 2012; Trager and Vavrek, 2011; Davies and Johns, 2013; Brutger, 2014), there has been relatively little attention paid to the relevant characteristics of audiences themselves. This asymmetry stems partially from the prevailing tendency of IR scholars to focus predominantly on situational features rather than dispositional characteristics, partially because formal approaches to audience cost theory tend to assume a homogenous audience for purposes of analytic tractability, and also because those experimental studies that have looked for heterogeneous treatment effects have tended to find that the logic of audience costs is impervious to the types of individual differences that routinely play an important role in public opinion about foreign policy more generally.2 The exception to this rule has been work on political identification and engagement, although the findings have been mixed (Tomz, 2007; Levendusky and Horowitz, 2012; Trager and Vavrek, 2011; Davies and Johns, 2013). In short, the experimental literature on audience costs has tended to interpret the large audience cost treatment effects as a sign that publics care about consistency, and the general lack of heterogeneous treatment effects as a sign that the logic of audience costs manifests itself across different subsets of the population.

2On the importance of dispositional characteristics in public opinion towards the use of force, see Hermann, Tetlock, and Visser, 1999; Kertzer and McGraw, 2012.
3.3 Audience Costs as a Double-Barreled Treatment Effect

We believe political scientists should exercise caution before drawing these conclusions, because there are more to audience costs than meets the eye. One of the fundamental tasks in designing and analyzing experiments is establishing the relevant counterfactual, selecting the control group to be compared to the treatment (Morgan and Winship, 2007). In audience cost theory, as Figure 3.1 illustrates, both the treatment and the control are clearly specified: the treatment is the President threatening force and failing to follow through, while the control is the President deciding to stay out of the conflict in the first place. In this respect, however, the treatment and the control differ from one another in two different ways: both in terms of whether the President says one thing and does another, but also whether the President threatens to use force in the first place.

![Figure 3.2: Experimental Design and Revised Crisis-Bargaining Model](image)

The left-hand side of the figure represents the structure of the experimental design as a 2x2 with the expected value in each cell written in experimental notation; the right-side side represents the same structure, this time as a game tree, with the expected payoffs for player 1 presented under each terminal node.

Since experimentalists and crisis bargaining scholars often speak subtly different languages, we make this point in two different ways in Figure 3.2. The left-hand side of Figure...
models leaders’ decisions in audience cost theory as experimentalists would, with a $2 \times 2$, in which the vertical axis refers to leaders’ initial decision about whether to threaten the use of force or not (which we refer to here as $\alpha$), and the horizontal axis depicts the subsequent decision about whether to back down or follow through on the initial decision (which we refer to here as $\beta$). The top-left cell (with expected value $\mu_{11}$) represents the control group from audience cost experiments, in which a leader decides to stay out, while the bottom-right cell (with expected value $\mu_{22}$) represents the classic audience cost treatment, in which a leader threatens the use of force, and fails to follow through. Since the audience cost treatment effect is defined as $\mu_{22} - \mu_{11}$ — the average difference in responses across the main diagonal in the $2 \times 2$ — the treatment and control differ from one another in both $\beta$ and $\alpha$, both the leaders’ deeds, and the leaders’ words: not just whether the leader follows through, but also whether she threatens to use force in the first place. In this sense, the audience cost treatment effect is “double-barreled.”

To be clear, our claim is not that audience cost experiments are unfaithful to audience costs, but rather that because the model the experiments are testing involves comparing outcomes at two different levels of the game tree, both the experiments and the model they are based on inevitably employ an “indissoluble double stimulus” (Converse and Presser 1986, 13). Indeed, we express the same argument using a revised crisis bargaining model in the right-hand side of Figure 3.2, which we turn to later. More important for our purposes is that this double-barreled treatment effect has implications both for how we understand the logic of audience costs, and how we test whether the same logic of audience costs manifests itself across different segments of the population. First, there are a variety of reasons why members of the public might disapprove of the President threatening force, ranging from doves who believe that military options only make problems worse (Holsti 1979), to Jeffer-
sonians who would rather the US focus on its own problems (Mead, 2002), to realists who see the President getting involved in a situation where the national interest is not at stake as a sign of a lack of competence (Kertzer and McGraw, 2012). If the total audience cost treatment effect represents both the impact of inconsistency and the impact of threatening force, there is no longer a unitary logic of audience costs. Different audiences can punish leaders for very different rationales, and we may wish to explore how much of the total audience cost comes from concerns about inconsistency versus concerns about the use of force. Second, if audiences punish leaders for different reasons, simply looking at whether the size of the audience cost treatment effect varies across different segments of the population obscures the possibility that the magnitude of the treatment effect might remain the same even if the mechanism behind it varies dramatically.

To disentangle these logics, we adopt a mid-way point between the traditional audience cost experiment and a fully-crossed $2 \times 2$ factorial, illustrated by the left-hand panel of Figure 3.2: we supplement the classic two-condition audience cost experiment with a third experimental condition in which the leader threatens the use of force and follows through — the bottom-left cell (with expected value $\mu_{21}$) in the $2 \times 2$ factorial. This simple modification is important, because as Figure 3.2 shows, this “Engage” experimental condition is a middle ground between the “Stay Out” condition and the “Not Engage” condition: it differs from the “Not Engage” condition in terms of whether the leader is consistent (the presence of the threat of force is held constant), and differs from the “Stay Out” condition in terms of whether the leader threatens the use of force (consistency is held constant). The absence of the fourth cell means we cannot analyze the experimental results as if the experimental results

---

4 There are a number of alternative approaches that could be used to disentangle these logics, including the use of open-ended responses, which we discuss along with a number of alternative designs in Appendix §3. In an innovative exploration of crisis bargaining models, Trager and Vavrek (2011) also measure public approval at the “Engage” node of the model, but because their experiment was designed to answer a different set of questions, they vary military outcomes rather than holding them constant, which precludes us from being able to decompose the audience cost treatment into inconsistency costs and belligerence costs in their design, for reasons we explain below.
design were a traditional $2 \times 2$ factorial, since the two factors are not fully crossed. However, since the “Engage” experimental condition ($\mu_{21}$) is a middle ground between the “Stay Out” condition ($\mu_{11}$) and the “Not Engage” condition ($\mu_{22}$), the total audience cost treatment effect $\mu_{22} - \mu_{11}$ can be re-expressed as the sum of two separate treatment effects:

$$\mu_{22} - \mu_{11} = (\mu_{21} - \mu_{11}) + (\mu_{22} - \mu_{21}) \quad (3.1)$$

To make this analytical move, we need to ensure the experiment holds outcomes constant across conditions: after all, we cannot capture how much the audience dislikes the threat of force if staying out results in splendid isolation while going in leads to a costly quagmire. More formally, we need to isolate the treatment effects from the audience’s perception of payoffs. In the classic crisis bargaining model displayed in Figure 3.1, actors make decisions based on the expected payoffs at the terminal nodes of the game tree. The initiating state is assumed to be fighting over some policy or good with value $v$ (usually normalized to a value of 1), and in the event it makes a threat and follows through, its expected payoff is $w_1$, a war payoff that incorporates the state’s belief about the probability of winning the good of value $v$ and some expected cost paid for engaging in military action. In observed crises, it would be nearly impossible to disentangle audiences’ approval for the value of the policy outcome from the strategy used to achieve it. The advantage of experiments, however, is that we can ensure participants are presented with an identical outcome (worth $v$) regardless of whether the President stays out, engages, or does not engage, such that changes in public approval can be traced entirely to the leader’s strategy. We can therefore revise the payoffs from the classic crisis bargaining model in Figure 3.1 in two ways: first, denoting these identical outcomes across the nodes of the game tree, and second, decomposing audience costs by adding a new payoff, $-f$, reflecting the possibility that audiences punish leaders for threatening force rather than just for being inconsistent.
Six points are important to note about this payoff structure, depicted in the game tree in the right-hand side of Figure 3.2. First, incorporating the $-f$ term renders transparent the double-barreled nature of audience cost models, since the audience cost is now $-f - a$: the sum of two different treatment effects. Second although the notation is different, the game tree in the right-hand side of Figure 3.2 corresponds exactly to the experimental design in the left-hand side. Comparing approval at the Stay Out and Engage nodes of the tree isolates $-f$, the belligerence cost — the simple effect of threatening force ($\mu_{21} - \mu_{11}$ from the $2 \times 2$), while comparing approval between the Engage and Not Engage nodes produces $-a$, the inconsistency cost — the simple effect of inconsistency ($\mu_{22} - \mu_{21}$). Third, it is worth re-emphasizing that $-f$ and $-a$ are simple effects rather than main effects; we refer interested readers to Appendix §4 for a discussion of why the former are the quantities of interest in audience cost models, rather than the latter.

Fourth, if threatening force has no impact on respondents’ evaluations of the President (that is, $-f = 0$), the conventional wisdom about audience costs will be correct, in that we should interpret the audience cost treatment effect as being driven by concerns about inconsistency, and all of the various mechanisms (national honor, leaders’ competence, etc.) it implicates, since if $-f = 0$, then $-f - a = -a$. If, on the other hand, $-f \neq 0$, at least some of the audience cost treatment effect is driven by concerns about the threat of force, and there is no longer a unitary logic of audience costs. This point can be understand both as a narrower methodological claim, and as a broader substantive one. On the one hand, it is clear that incorporating a double-barreled design means that classical audience cost experiments offer a systematically biased estimate of how much people care about inconsistency. This point is subtly different from critiques raised by Snyder and Borghard (2011), Chaudoin (2014b) and others: our concern is not so much that people care less about inconsistency than audience cost theory assumes, but that because of their double-barreled design, these traditional experiments cannot tell us how much people care about inconsistency at all.
On the other hand, however, these classical experiments are inheriting the assumptions of the crisis bargaining model they are based on; the potential for belligerence costs thus raises interesting implications for crisis bargaining scholars more generally, with important ramifications for how leaders send signals in international crises.

Fifth, audience costs have traditionally been understood as as a way for leaders to tie their hands: unlike sunk-cost signals that are immediately costly for leaders to carry out, but should not affect their future behavior, hands-tying incentivizes future behavior by producing costs that are only paid if the leader does not follow through \( \text{[Fearon 1997]} \). If audiences primarily punish leaders for being inconsistent, public threats should be informative signals for crisis bargaining and function in the manner consistent with audience cost theory, since leaders will pay the cost only if they back down on the threat. If, however, audiences punish leaders for making threats, then public escalation generates sunk costs, since the cost is paid when the threat is made, thereby invoking an alternative signaling mechanism and undermining the prevailing understanding of the signaling value of audience costs.

Finally, if audience costs have two distinct logics, we should be able to estimate these quantities of interest across different types of respondents (hawks versus doves, nationalists who see their country as superior to others, versus cosmopolitans who identify with the broader global community, etc.) and observe the extent to which the logic of audience costs vary. If we find that some audiences punish leaders for making threats and others punish inconsistency, the signaling value of public escalation will be even murkier, as different audiences impose costs for different reasons. This means signaling in crises is contingent on both the president’s strategy and the relevant audience. To effectively interpret the signaling value of public escalation a foreign adversary would have to evaluate the relative importance of different audiences to the leader, which would make public escalation a noisier signal in international crises than it is often considered.
3.4 Bringing the Audience Back In

One of the advantages of decomposing audience costs into concerns about inconsistency versus concerns about the threat of force is that it opens up the possibility of integrating audience cost theory with a large body of research in political psychology and public opinion about foreign policy that explores the dispositional underpinnings of foreign policy attitudes. As noted above, existing audience cost experiments have largely shied away from investigating how demands for audience costs vary with dispositional characteristics of audiences, but those investigations that have taken place have tended to search for dispositional effects by testing whether the size of the audience cost treatment effect varies across subgroups (e.g., Tomz [2007] Davies and Johns [2013]). However, if audience cost treatment effects actually contain two different quantities of interest, the logic of audience costs might vary dramatically between hawks and doves, say, even if the magnitude of the audience cost does not significantly differ. There is no shortage of dispositional characteristics we could investigate, but for theoretical reasons we focus on four characteristics here, each one corresponding to a slightly different logic.

First is militant assertiveness. Although, as Weeks (2008, 43) reminds us, “the “audience cost” does not arise because domestic audiences disagree with their leaders’ policy,” if audience costs actually involve two different costs — one reflecting inconsistency, and the other reflecting the threat of force — then understanding variation in audience costs requires us to model substantive policy preferences as well (Snyder and Borghard [2011] Levy [2012] Chaudoin [2014b]). A large volume of research in public opinion on foreign policy tells us that the use of force is an issue about which people’s attitudes vary systematically (Holsti [1979] Chittick, Billingsley, and Travis [1995] Murray [1996] Herrmann, Tetlock, and Visser [1999] Rathbun 2007 Gelpi, Feaver, and Reifler [2009]). Beliefs about the desirability and effectiveness of the use of force form a major part of all of the classic models in the American foreign policy attitudes literature, even if this construct takes a number of different names:
Wittkopf (1990) (and later, Holsti and Rosenau 1990) refer to “militant internationalism”, Hurwitz and Peffley (1987) to “militarism”, and Herrmann, Tetlock, and Visser (1999) to “militant assertiveness.” Similar constructs have been employed in analyses of the structure of foreign policy attitudes in countries outside the United States as well (Hurwitz, Peffley, and Seligson, 1993; Bjereld and Ekengren, 1999; Reiffer, Scotto, and Clarke, 2011). Audiences should care far less about inconsistency if they disapproved of the policy the leader is backing down from in the first place: thus, since hawks and doves differ in their beliefs about whether governments should use military means to achieve their foreign policy ends, we expect that individuals who are low in militant assertiveness will demand audience costs primarily to punish the President for threatening to use force, but will be less concerned about inconsistency, while those who are high in militant assertiveness should support the threat of force and thus demand audience costs for failing to use it.

Second is international trust. According to one of the classic interpretations of audience cost models, audiences dislike inconsistency because it sullies the country’s reputation for resolve and weakens its future bargaining position (Guisinger and Smith, 2002; Tomz, 2007). In this context, like in many rationalist models, reputation and trust are inextricably linked, since institutional environments that incentivize the building of the former facilitate the emergence of the latter (e.g., Milgrom, North, and Weingast, 1990; Kydd, 2005). Yet as Rathbun (2009, 2011) argues, this type of “strategic trust” based on incentive structures is at odds with much of the psychological and sociological literature on trust, which also understands trust as moralistic or fiduciary, based on beliefs about the character of general or specific others (Cook, 2001; Uslaner, 2002; Hoffman, 2002; Hardin, 2006). Whereas strategic trust is situationally derived, moralistic trust is dispositionally varying: some people are generally more trusting than others, endowed with a sunnier view of human nature that “allows one to go beyond incentives” (Mercer, 2010, 6), be more prone to cooperation, and less concerned about exploitation (Kuhlman and Wimberley, 1976; Wrightsman, 1991).
In this sense, it should not be surprising that variation in trust has been found to predict attitudes towards a wide range of foreign policy issues, from the use of force to support for free trade (Brewer and Steenbergen, 2002; Chanley, 2002; Brewer, 2004; Binning, 2007; Kaltenthaler and Miller, 2013). Individuals who display less international trust, who are more cynical about the prospects of exploitation by others and thus should be most concerned about maintaining reputations for resolve, should demand audience costs primarily to punish the President for displaying inconsistency, and should hardly be bothered by the President’s threat of force. In contrast, individuals with high levels of international trust, who are more sanguine about the motivations of other countries and should thus be less preoccupied with maintaining reputations for resolve, should be more likely to punish Presidents for making threats than for inconsistency.

Third is nationalism. As Tomz (2007, 835) notes, many reputation-based accounts of audience cost models offer what one might call an instrumental or consequentialist (March and Olsen, 1998) understanding of audience costs, in which audiences punish inconsistency because of the negative consequences they believe result from it. However, we can also think of audience cost mechanisms as following not an instrumental concern about reputation or credibility, but a moral or normative concern about national honor. Thus, another key audience characteristic relevant for our purposes is what Chittick, Billingsley, and Travis (1995) call “identity”, Rathbun (2007) calls “community”, and what social psychologists would call “ingroup identification”: the extent to which individuals are nationalists with

---

5 In Uslaner’s 2002 framework, moralized trust can be particularized (“I trust you”) or generalized (“I trust”); many of the kinds of trust invoked in political science – social trust, international trust, and so on — are of the generalized form, so we focus on it here.

6 This is clearly the case for audience cost models that discuss negative bargaining reputations, but is also compatible with models in which audiences use inconsistency as a signal of a leader’s lack of competence (?): why would backing down be a sign of incompetence if it had positive consequences?

7 On the distinction between reputation and honor, see Schelling 1966, 124, O’Neill 1999, Dafoe, Renshon, and Huth 2014. Indeed, Fearon (1994) explicitly explains audience costs in terms of “engaging the national honor”, although the historical cases he discusses can be understood as invoking honor for both the sake of a logic of consequences, and for a logic of appropriateness. On moral underpinnings of foreign policy attitudes, see Kertzer et al. (2014).
deep identifications with their nation (and thus more willing to use force on its behalf), or cosmopolitans who identify with the international community as a whole. Building off of work on social identity theory — which distinguishes between ingroup favoritism and outgroup derogation (Brewer, 1999) — IR scholars who have studied ingroup identification tend to distinguish between national attachment versus national chauvinism, the latter of which examines individuals’ tendencies to not only identify with their nation, but also view it as superior to others (Herrmann, Isernia, and Segatti, 2009). National chauvinists, who have little qualms about using force on their ingroup’s behalf, should punish the President not for threatening the use of force, but for violating the national honor by being inconsistent. In contrast, individuals who are low in national chauvinism, who are less likely to believe in the inherent inferiority of other countries and thus less willing to resort to force, should be relatively less sensitive to inconsistency and more to the threat of force.

Fourth is political ideology — the broader belief system that structures how people think about politics, typically represented as a continuum with liberals on the left, and conservatives on the right (Converse, 1964; Jost et al., 2003b; Feldman, 2003). Although characteristics like international trust and militant assertiveness are helpful in offering clearly-specified microfoundations for public opinion, holistic constructs like ideology are also useful because they encapsulate these more specific traits within a broader framework. A plethora of psychological research emphasizes the motivational underpinnings of political ideology, positing that individuals are attracted to liberal or conservative belief systems to fulfill discrete psychological needs (Jost et al., 2003b, 2007). Conservatives are higher not just in needs for order, structure, and cognitive closure (Webster and Kruglanski, 1994) that should make them less tolerant of inconsistency, but also in authoritarianism, social dominance, and system justification (Altemeyer, 1998; Sidanius and Pratto, 2001; Duriez and Van Hiel, 2002; Jost, Banaji, and Nosek, 2004), which should make them more acceptant of the use of force.

---

8See also Schatz, Staub, and Lavine 1999 on “blind” versus “constructive” patriotism, and Nincic and Ramos 2012 on “uncritical” versus “critical” patriotism.
They are more likely to subscribe to “competitive jungle” and “dangerous world” beliefs (Duckitt et al., 2002), that should enhance their concern about reputations for resolve, while serving as license for the use of force. These existential and epistemic motives cluster together, but point in the same direction: we should expect conservative audiences to punish leaders for inconsistency, while liberal audiences will be relatively more concerned about the use of force.

Two additional considerations are worth keeping in mind. First, each of our theoretical expectations outlined above suggest that at high and low levels of these dispositional characteristics, audience costs may follow a logic of substitutable rather than conjunctive causation (Braumoeller, 2003): some segments of the audience will demand inconsistency costs and others belligerence costs, rather than punishing for both reasons simultaneously. Second, needless to say, there are a far larger number of audience characteristics we could potentially examine than the four traits we discuss above. To the extent that we find evidence for heterogeneous treatment effects across these four traits, it is reasonable to assume that we would find further evidence of heterogeneity across other audience characteristics as well.

3.5 Method

To test our theoretical framework we fielded an online survey experiment in the spring of 2014 on a national American sample of 942 registered voters recruited by Survey Sampling International (SSI)\textsuperscript{9}. The study includes an audience cost experiment (described in detail below), as well as a dispositional questionnaire measuring the series of individual-level characteristics outlined above, allowing us to examine the extent to which different types of

\textsuperscript{9}For examples of recent political science survey experiments fielded on SSI samples, see Malhotra, Margalit, and Mo (2013); Berinsky, Margolis, and Sances (2014). See Appendix §2 for sample characteristics and more details about the sampling methodology. We also replicated our results using an online sample of 1345 American adults recruited on Amazon Mechanical Turk.
members of the public demand audience costs for different types of logics. Since the purpose of the study is to challenge the conventional interpretation of audience costs, the structure of the audience cost experiment borrows heavily from the classic audience cost experiments described above. In the scenario, respondents were first presented with an introduction that read:

The following questions are about U.S. relations with other countries around the world. You will read about a situation our country has faced many times in the past and will probably face again. Different leaders have handled the situation in different ways. We will describe one approach U.S. leaders have taken, and ask whether you approve or disapprove.

Respondents were then presented with information about a hypothetical international crisis, which, like most audience cost experiments, evokes the classic “repel an invader” experiment first introduced in IR by Herrmann, Tetlock, and Visser (1999), in which a foreign country sends its military to take over part of a neighboring country. Following existing audience costs experiments, we deliberately chose not to identify the foreign countries in the scenario so as to avoid introducing country-specific confounding, and to more directly speak to the existing literature. After participants read that “A country sent its military to take over a territorial region in a neighboring country,” respondents were randomly assigned to one of three different experimental conditions, reflecting the three different cells in Figure 3.2. In the “Stay Out” condition, participants were told that the President announced that “the United States would stay out of the conflict,” whereupon the attacking country continued to invade. In the “Not Engage” condition, participants were told that the President announced that “if the attacking country continued to invade, the United States military would immediately engage and attempt to push out the attacking country”; the attacking country continued to invade, but the President did not engage.
Thus, as in all audience cost experiments, the “Stay Out” condition and the “Not Engage” condition vary both (i) whether the President threatens the use of force, and (ii) whether the President says one thing but does another. We therefore introduce a third experimental condition as a middle ground between these two treatments: in the “Engage” condition, the President announces that “if the attacking country continued to invade, the United States military would immediately engage and attempt to push out the attacking country”; the attacking country continued to invade, and the President ordered the US military to engage. The treatment thus differs from each of its counterparts in an important manner: it differs from the “Not Engage” treatment in terms of whether the President displays inconsistent behavior, and differs from the “Stay Out” condition in terms of whether the President’s initial commitment is to stay out or to use force. To control for policy outcomes, across all three treatments the final result to the United States is the same: the conflict ends with the hostile state gaining 20 percent of the contested territory regardless of whether the President goes in or stays out, and because the war payoff generally includes an additional cost of fighting, we specify in the “Engage” condition that the U.S. experienced zero casualties in the conflict. Doing so not only minimizes the perceived cost of war, but also serves as a more conservative test; as we show in Appendix §5, replicating the experiment without specifying the absence of casualties in the “Engage” condition produces larger estimates of the cost of threatening force. In all three conditions, we measure our dependent variable of interest by asking participants the extent to which they approved or disapproved of the way the President handled the situation, which yielded a seven point approval scale.

We also include a questionnaire measuring a series of dispositional characteristics borrowed from the political psychology and public opinion about foreign policy literatures: we measure participants’ militant assertiveness using items borrowed from Herrmann, Tetlock, and Visser (1999), national chauvinism using items from Herrmann, Isernia, and Segatti (2009), international trust from Brewer (2004), and political ideology using the standard
measure from the American National Election Studies (ANES). We also measure a variety of standard demographic characteristics including gender and education.10

3.6 Results

We present our results in two phases. First, we simply look at our treatment effects, decomposing audience costs into concerns about inconsistency versus concerns about threatening force. Second, we use the individual differences measured in our dispositional questionnaire to show how different kinds of respondents weigh these two rationales differently, such that audience costs can be demanded for very different reasons by different audiences.

3.6.1 Average Treatment Effects

As an initial examination of the results, we estimate the audience cost treatment effect \( \mu_{22} - \mu_{11} \), the average difference in support for the President when the President threatens force and backs down, compared to the President choosing to stay out of the crisis. As is the case in all audience cost experiments, the audience cost treatment is negative and statistically significant: approval for the President is an average of 1.35 points lower on a seven point scale (95% bootstrapped CI: -1.71, -0.98) when the President backs down on a threat than when she follows through on staying out in the first place. However, as previously discussed, this treatment effect is double-barreled: it contains not only the costs produced from backing down, but also those incurred from threatening to get involved in the first place. When we decompose this treatment effect into its constituent parts, we can see that our participants indeed punish the leader for being inconsistent, in that \( \mu_{22} - \mu_{21} \) — the average difference in support for the President when the President threatens force and backs down versus threatens force and follows through — is negative and statistically significant.  

10In this study, we administered the dispositional and demographic batteries prior to the foreign policy scenario; to test for downstream effects, in other versions of the study we have also administered the two parts in the reverse order, and find the results hold.
significant: on average, approval for the President is 0.91 points lower (95% bootstrapped CI: -1.31, -0.47) when the President threatens force and backs down than when she threatens force and follows through. However, \( \mu_{21} - \mu_{11} \) — the average difference in support for the President when the President follows through on a threat of force versus follows through on a promise to stay out — is also negative and statistically significant: on average, approval for the President is 0.45 points lower (95% bootstrapped CI: -0.90, -0.02) when the President follows through on a threat to use force than on a pledge to stay out.\(^{11}\)

Two points are worth noting here. First, note that the total audience cost treatment effect is a sum of inconsistency costs and belligerence costs; this is not simply a coincidence of the data, but axiomatically true, since by definition, \((\mu_{21} - \mu_{11}) + (\mu_{22} - \mu_{21}) = \mu_{22} - \mu_{11}\). Second, as Gartzke and Lupu (2012) point out, audience costs are about a specific mechanism, rather than just an effect. Yet these results show us that audience cost experiments are measuring more than just concerns about inconsistency: 67% of the audience cost treatment effect in our sample comes from an inconsistency cost, but 33% comes from a belligerence cost. Third, although classical audience cost studies have tended to treat the public as homogenous, the public opinion literature has tended to find considerable heterogeneity both in terms of what people think about foreign affairs, and how they think about it. As we see below, the presence of these two different forms of punishment offers one potential reason why it has been surprisingly difficult to find evidence of heterogeneous treatment effects.

### 3.6.2 Heterogenous Logics of Audience Costs

The conventional way of investigating heterogeneity with audience costs would be to estimate a series of statistical models in which we interact the audience cost treatment with a number of dispositional characteristics, to see whether certain types of people demand larger audience

\(^{11}\)We can also examine belligerence and inconsistency costs by measuring changes in the percentage of the audience that disapproves of the President’s handling of the situation (Tomz, 2007, 835). For the full sample, inconsistency increases disapproval by 21 percentage points, whereas belligerence increases disapproval by 8 percentage points. For additional results in percentage point form, see Appendix §7.
costs than others (e.g., Tomz, 2007; Levendusky and Horowitz, 2012; Davies and Johns, 2013). Our interest here, however, is somewhat different, since we wish to disentangle the relative importance of each of these two logics. In lieu of complex statistical models, we keep the analyses simple by estimating a series of subgroup analyses, where we split the sample in two (separating hawks from doves, for example), and decompose the audience cost in each subgroup, estimating what fraction of the audience cost effect comes from inconsistency costs ($\frac{\mu_{22} - \mu_{21}}{\mu_{22} - \mu_{11}}$), versus belligerence costs ($\frac{\mu_{21} - \mu_{11}}{\mu_{22} - \mu_{11}}$). In Appendix §6, we replicate these analyses in a regression context that controls for all of the dispositional variables simultaneously, and show the substantive results remain the same.

Table 3.1: Decomposing Audience Costs: Threatening Force Versus Inconsistency

<table>
<thead>
<tr>
<th>Audience Cost</th>
<th>Belligerence Cost</th>
<th>Inconsistency Cost</th>
<th>Belligerence Cost Fraction</th>
<th>Inconsistency Cost Fraction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full Sample</td>
<td>-1.353</td>
<td>-0.447</td>
<td>-0.906</td>
<td>0.326</td>
</tr>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
</tr>
<tr>
<td>Low Militant</td>
<td>-1.770</td>
<td>-1.754</td>
<td>-0.017</td>
<td>1.016</td>
</tr>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
</tr>
<tr>
<td>High Militant</td>
<td>-1.284</td>
<td>0.194</td>
<td>-1.478</td>
<td>-0.225</td>
</tr>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
</tr>
<tr>
<td>Low Int Trust</td>
<td>-1.304</td>
<td>-0.135</td>
<td>-1.169</td>
<td>0.090</td>
</tr>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
</tr>
<tr>
<td>High Int Trust</td>
<td>-1.462</td>
<td>-1.548</td>
<td>0.086</td>
<td>1.173</td>
</tr>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
</tr>
<tr>
<td>Low Nat Chauv</td>
<td>-1.122</td>
<td>-0.522</td>
<td>-0.599</td>
<td>0.459</td>
</tr>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
</tr>
<tr>
<td>High Nat Chauv</td>
<td>-1.554</td>
<td>-0.535</td>
<td>-1.019</td>
<td>0.337</td>
</tr>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
</tr>
<tr>
<td>Liberals</td>
<td>-1.058</td>
<td>-0.870</td>
<td>-0.188</td>
<td>0.882</td>
</tr>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
</tr>
<tr>
<td>Conservatives</td>
<td>-1.516</td>
<td>-0.386</td>
<td>-1.130</td>
<td>0.224</td>
</tr>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
</tr>
</tbody>
</table>

Note: Columns 1-3 depict average treatment effects calculated from 2000 bootstraps, shown with both experimental and crisis bargaining notation; note that the threaten force cost (2) and inconsistency cost (3) add up to the total audience cost (1). Columns 4-5 calculate the fraction of the total audience cost stemming from threatening force (4), versus from inconsistency (5). Note that these two fractions add up to 1. For more on these quantities of interest, please refer to Figure 3.2.

We present the results both numerically in Table 3.1 and visually using density plots in Figure 3.3. The first three columns of Table 3.1 display the three treatment effects (audience

\[ \mu_{22} - \mu_{11} \] versus belligerence costs \[ \mu_{21} - \mu_{11} \]. For all subgroup analyses, we define low and high levels of each dispositional characteristic using the interquartile range, thus comparing individuals in the bottom and top 25%, but this method obviously generalizes to the specification of other cutpoints as well.
Figure 3.3: The multiple logics of audience costs

Note: For both the full sample (panel a) and the series of subsamples depicted in panels b-h, Figure 3.3 decomposes the audience cost treatment effect $\mu_{22} - \mu_{11}$ into two separate treatment effects: concerns about threatening force ($\mu_{21} - \mu_{11}$) in dark grey and concerns about inconsistency ($\mu_{22} - \mu_{21}$) in light grey, each one expressed as a fraction of the total audience cost treatment effect, and illustrated as density distributions. The x-axes are bounded at (-0.4,1.4) for ease of interpretability. These density plots shows that audience costs display multiple logics: hawks (panel c) and doves (panel b), for example, demand audience costs for very different reasons.
costs, belligerence costs, and inconsistency costs), while the two right-hand columns decom-pose the audience cost effect, showing our two main quantities of interest: the fraction of the audience cost effect that comes belligerence costs, versus from inconsistency costs. These two quantities of interest are displayed visually in Figure 3.3 in which the fraction of audience costs stemming from inconsistency is illustrated in light grey, and the fraction of audience costs stemming from concerns about the use of force is displayed in dark grey, the probability distributions are derived from 2000 bootstraps to provide estimates of uncertainty. Thus, the first row of the table and panel a of Figure 3.3 show that as discussed above, 67% of the audience cost effect in our full sample comes from inconsistency costs, while 33% comes from belligerence costs. While our results replicate existing studies that find significant audience costs, it is also clear that we risk seriously misunderstanding the logic underpinning why audiences punish leaders in domestic crises if we attribute the presence of audience costs solely to the public’s punishment of inconsistency. The question is whether this pattern replicates when we start disaggregating the audience.

The first dispositional characteristic we look at is militant assertiveness, which illustrates how policy preferences drive audience costs. When we compare those participants whose militant assertiveness scores in the dispositional survey placed them in the bottom 25% (panel b) versus those in the top 25% (panel c), we can see that doves and hawks demand audience costs for fundamentally different reasons. We find no evidence that respondents who are low in militant assertiveness are turned off by inconsistency — indeed, the inconsistency treatment effect is not statistically significant from zero (p < 0.49) — but are heavily concerned about the leader threatening force (p < 0.00), which is responsible for the entire cost they impose. In contrast, although those respondents who are high in militant assertiveness also demand an audience cost, it is entirely driven by the leader being inconsistent. Two points are noteworthy here. First, although the idea that doves and hawks should have different preferences about foreign policy might seem unsurprising, this finding challenges all previous
experimental work on audience costs we are aware of, which has found hawks and doves to be relatively homogenous in how they demand audience costs, and thus assumes there to be very little variation in how much people care about leaders who display inconsistent behavior. Second, simply looking at the magnitude of the audience cost effect between doves and hawks would be misleading: doves in our sample actually demand larger audience costs than hawks do, but this effect is a consequence of hawks rewarding the President for threatening force; in our sample, focusing solely on the magnitude of audience costs thus not only overestimates how much doves care about inconsistency but also underestimates how much hawks care about inconsistency by 15%.

Second, we turn to international trust, which as noted above, offers one way of testing the reputational mechanism posited to drive audience costs. Since individuals who are lower in international trust are more cynical about their country’s exploitation by others and thus have the most reason to be preoccupied with maintaining reputations for resolve, we would expect them to demand audience costs primarily to punish the President for displaying inconsistency and thereby harming the country’s bargaining reputation. In contrast, individuals with higher levels of international trust, guided by sunnier views of the motives of other countries, should be less concerned about inconsistency and more about the threat of force. Sure enough, 91% of the audience cost for respondents low in international trust is driven by concerns about inconsistency (panel d of Figure 3.3), while respondents high in international trust (panel e) demand audience costs entirely because of the threat of force. On the one hand, then, these findings offer evidence in favor of reputational understandings of audience costs, but on the other, it also reminds us that not all audiences are concerned about reputational costs.

Third, we turn to national chauvinism, which ties into accounts of audience costs that stress violations of the “national honor.” We articulated above why nationalists, who are the most likely to embrace notions of national honor, should be more likely to punish the
President for inconsistency, while cosmopolitans, less likely to see the US as superior to other nations, should care less about inconsistency, and more about the threat of force. Panels f and g of Figure 3.3 offer some support for this hypothesis: nationalists who see the United States as far superior to other nations strongly punish the President for being inconsistent (panel g), while relatively “cosmopolitan” respondents low in national chauvinism (panel f) display relatively weak effects, split between the two logics (p < 0.10 for inconsistency costs, p < 0.14 for belligerence costs). However, unlike with militant assertiveness and international trust, the difference in the fraction of audience costs driven by inconsistency across the two subgroups is not statistically significant (p < 0.381). At least as proxied by national chauvinism, then, we find less evidence for national honor-based explanations of audience costs than for the other mechanisms tested here.

Finally, panels h and i of Figure 3.3 looks at political ideology, disaggregating audience costs among self-reported liberals versus conservatives. As expected, audience costs among liberals are driven primarily by belligerence costs, while a full 78% of conservatives’ audience cost comes from inconsistency. Once again, then, we find signs of heterogeneous logics of audience costs, showing that classical audience cost experiments are painting a somewhat distorted picture of the dynamics of public opinion in foreign policy crises: both liberals and conservatives demand audience costs, but for very different reasons.

3.7 Implications

These dual forms of audience costs provide a series of substantive insights into the value of public threats as signals in international crises, four of which we briefly touch on here.

First, audience costs include a sunk cost. Until now, audience cost scholars have focused on signaling through hands-tying, in that threats were understood to generate costs that were only paid if the leader chose not to follow through. In contrast, we find that leaders also pay a sunk belligerence cost when they make threats: in a population composed of
both doves and hawks, for example, the President loses the support of the doves as soon as she issues the threat of force, and does not regain their approval regardless of whether she follows through or not. Having already made a threat, her level of support now depends on the approval of hawks, whose adamant aversion to inconsistency means that she can either follow through on her commitment, or back down and alienate both segments of population. The leader is thus faced with an initial decision of whether to pay a belligerence cost when the threat is made, while also generating additional potential costs that will only be paid at a later date, depending on her decision to follow through or not.

Second, threats can screen: even when an audience does not punish their leader for failing to follow through on a threat, the public threat itself generates a domestic belligerence cost that directly affects the leader’s approval (and payoff in the crisis bargaining model), which can thus be used as a costly signaling mechanism, enabling foreign adversaries to learn about the threatening state’s level of resolve. Since a leader would only initiate a threat when the expected value of the prize is greater than the expected cost of the threat, initiating a threat acts as a credible screening device. Only leaders who place a sufficiently high value on the contested prize would be willing to publicly threaten, which means foreign leaders can update their beliefs when public threats are observed. In this sense, although IR scholars have generally understood diplomatic communication to be of limited value in the absence of traditional audience costs (Fearon, 1995; Guisinger and Smith, 2002), our findings offer additional support in favor of Sartori’s (2002) and Ramsay’s (2011) arguments that talk in IR is less cheap than is often claimed. This point also illustrates an important difference from other work critiquing audience cost theory: Snyder and Borghard (2011), for example, suggest that empty threats are less costly than audience cost theory assumes.

---

13 Any strategy where the threatening state pays a sunk cost greater than the value of the prize is strictly dominated (Fearon, 1997).
14 Baum (2004) also notes that leaders should evaluate the potential reward when making threats, but the key decision for the leader in his formulation is weighing the reward versus the risk of military defeat, which invokes a different calculation, since military defeat is a potential cost contingent on state power, as opposed to a sunk domestic cost.
while our findings about belligerence costs suggests that threats can be costly regardless of their emptiness.

Third, leaders have an incentive to pursue secret diplomacy. An expanding literature on the value of secret negotiations highlights the ability of closed-door communications to play an important role in bargaining (Kurizaki 2007; Ramsay 2011; Trager 2010; Yarhi-Milo 2013). Most notably, Yarhi-Milo (2013) outlines the informative role of secret communication when the revelation of the communication may have domestic costs. While Yarhi-Milo (2013) focuses on assurances, the same logic holds in the presence of belligerence costs. The threatening leader can avoid the sunk cost by conveying the threat in secret, yet the threat remains credible given that there is a risk the cost will be paid if the threat is later revealed.\footnote{It remains a task for future research to determine whether a leader may talk her way out of this type of punishment by providing a justification for her strategy, as Levendusky and Horowitz (2012) have argued can be done with traditional audience costs.} This suggests that there is no such thing as costless bluffing when it comes to public threats, and leaders have an incentive to exercise restraint when considering whether to initiate a threat under the eyes of a watchful public. As Sechser (2011) notes, public coercive threats are surprisingly scarce, and the presence of belligerence costs offers one potential explanation.

Fourth, the credibility of public threats depends on who the leader’s relevant audience is, and the results from the “heterogeneous logics” section illustrate a public sufficiently variegated that interpreting signals is all the more complicated (Jervis 1976). Not only do leaders’ key constituencies change over time thanks to the ebb and flow of domestic politics, but so too do the distributions of each of the audience characteristics we explored here, which tend to fluctuate in response to events at home and abroad (MacKuen, Erikson, and Stimson 1989; Keele 2007; Kertzer 2013). Thus, one would expect that as the composition of the public shifts, so too would the balance of considerations driving audience costs. Moreover, a host of psychological biases from the foreign policy decision-making literature suggests foreign leaders are thus likely to miscalculate the credibility of public threats in the face of mosaic...
audiences: belief perseverance and embedded images mean decision-makers will be slow to update their beliefs in response to these changes (Herrmann and Fischerkeller 1995), while the outgroup homogeneity effect makes it harder for decision-makers to pick up on divisions in rival foreign publics (Judd and Park 1988). The twin dynamics of belligerence and inconsistency costs thus offers further grist for Mercer’s (2012) concern about how cognitive limitations can contort the effects of signals, and invites additional skepticism about audience costs’ efficacy.

3.8 Conclusion

Given what Baum and Potter (2014) call the “cottage industry” of scholarship on audience cost theory, audience cost experiments have become one of the more popular experimental traditions in International Relations, despite facing a flurry of critiques. Some critics note that Presidents can escape punishment for inconsistency by pointing to new information, managing elite criticism, or by making concessions in private (Levendusky and Horowitz 2012; Saunders 2015; Brown and Marcum 2011). Other skeptics express concern that the logic of audience costs seems to better manifest itself in the lab rather than in crucial historical cases (Snyder and Borghard 2011; Trachtenburg 2012). Another group still suggests that citizens have preferences over policy choices and outcomes that may trump preferences about consistency (Snyder and Borghard 2011; Downs and Sechser 2012; Chaudoin 2014b). We argue here that even when we set these external critiques of audience cost theory aside, we need to be careful about the inferences we draw about the logic of audience costs, since the control and treatment groups differ in multiple ways, such that participants can punish both in response to inconsistency and for the leader threatening the use of force.

We presented the results here from a novel experimental design where we disentangle these two competing logics of why audiences punish. Our results show that there is no unitary logic of audience costs, and that some audiences impose inconsistency costs while
others impose belligerence costs. Bringing the audience back into our analysis gives us a richer understanding of the dynamics of public responses to crisis bargaining, and thus helps narrow the gulf between audience cost theory and the study of public opinion. The results presented here suggest that audience costs are not governed by a fundamentally different logic from that which governs public attitudes towards foreign policy issues more generally, in that the dispositional characteristics used by political psychologists and public opinion scholars to explain variations in foreign policy attitudes also explain differences in how people demand audience costs. These dual logics also raise important implications for crisis bargaining, explaining the rarity of public threats, and showing how their signaling value is contingent on the leader’s relevant constituency.
A.1 Summary of Key Variables:

- $I_G$ = Government initiates a WTO complaint.
- $L$ = Total litigation cost for a case.
- $L_j$ = Litigation cost paid by $j \in \{F, G\}$.
- $\tau_j(1)$ = Value of trade for $j$ with the barrier in place.
- $\tau_j(0)$ = Value of trade for $j$ with the barrier removed.
- $\theta_S$ : $Pr$(winning a strong case) $\sim U(0.5, 1.0)$, $\hat{\theta}_S = 0.75$
- $\theta_W$ : $Pr$(winning a weak case) $\sim U(0.0, 0.5)$, $\hat{\theta}_W = 0.25$
- $E_G$ = Externalities to the government of bringing the case.
A.2 Equilibrium Solution

A.2.1 Overview of Equilibrium Solution

The equilibrium consists of three regions of the parameter space and the equilibrium is unique within each:

For low cost cases, where \( L < \max\{\hat{\theta}_W(\tau_G(0)) + E_G + \hat{\theta}_W(\tau_F(0)), 0.5p + 0.5\} \), the firm plays the same strategy regardless of whether the case is strong or weak and the case is initiated by the government.

In the middle parameter space, where \( L \geq \max\{\hat{\theta}_W(\tau_G(0)) + E_G + \hat{\theta}_W(\tau_F(0)), 0.5p + 0.5\} \) and \( L \leq \hat{\theta}_S(\tau_G(0)) + E_G + \hat{\theta}_S(\tau_F(0)) \) the firm contributes if the case is strong, which leads to the government initiating the case, but will not contribute if the case is weak, in which case the government does not initiate.

Lastly, for high cost cases, where \( L > \hat{\theta}_S(\tau_G(0)) + E_G + \hat{\theta}_S(\tau_F(0)) \) the firm and government do not contribute and the case is not initiated, regardless of whether the case is strong or weak.

In the low and middle cost cases, if the firm chooses to contribute, it selects a litigation contribution from \( L_F > 0 \) that is the minimum amount to convince the government to initiate the case. The equilibrium contribution levels for both the firm and government for strong and weak cases (for \( P = 0.5 \) and \( E_G = 0 \)) are shown in Figure 2 of the paper and are proven in the following pages.
A.2.2 Proof of Equilibrium Solution:

For simplicity, $\tau_j(0) = 1$, $\tau_j(1) = 0$, and $E_G = 0$ in the following proofs.

**Low Cost Parameter Space:** $L < \max\{\widetilde{\theta}_W(\tau_G(0)) + E_G + \widetilde{\theta}_W(\tau_F(0)), .5p + .5\}$

1. Firm’s Strategy:

$$\sigma_F(m = s) = \sigma_F(m = w) : L_F = L - .5p + .25, \text{ if } L > .5p + .25$$

$$\sigma_F(m = s) = \sigma_F(m = w) : L_F = 0, \text{ if } L \leq .5p + .25$$

2. Government’s Beliefs:

$$Pr(\theta_S) = p$$

The government does not update its prior, since the firm is pooling and thus no information is conveyed.

3. Government’s Strategy: If $L_G \leq .5p + .25$ then:

$$\sigma_G = I_G, \text{ where } L_G = L - L_F$$

$$EU_G(I_G) = p(\widetilde{\theta}_S(\tau_G(0)) - L_G) + (1 - p)(\widetilde{\theta}_W(\tau_G(0)) - L_G) = .5p + .25 - L_G$$

$$EU_G(\neg I_G) = \widetilde{\theta}_W(\tau_G(1)) = 0$$

If $L_G > .5p + .25$ then:

$$\sigma_G = \neg I_G$$

$$EU_G(I_G|L_G = L) = p(\widetilde{\theta}_S(\tau_G(0)) - L_G) + (1 - p)(\widetilde{\theta}_W(\tau_G(0)) - L_G) = .5p + .25 - L_G < 0$$

$$EU_G(\neg I_G) = \widetilde{\theta}_W(\tau_G(1)) = 0$$

4. Sequential Rationality

We now consider whether the beliefs and strategies of the actors are sequentially rational.
If \( L \leq .5p + .25 \), the cost to initiate the case is low enough that the government initiates the case on its own and the firm has no incentive to deviate and contribute, since any additional contribution will be an unnecessary cost to the firm.

\[
EU_F(L_F = 0) = 0.25 > EU_F(L_F > 0) = .25 - L_F,
\]

(No incentive for the firm to deviate by contributing.)

If \( L > .5p + .25 \) (but still in the “low cost” space), the government will not initiate on its own, so the firm has an incentive to pay just enough to convince the government to initiate the case, up to the expected profit of winning a weak case. (No signaling occurs in this equilibrium - just a cost sharing of the litigation expenses.) If the firm deviated and chose to contribute less, the government would not initiate and the firm would lose the expected utility of initiating the case. If the firm paid any more than \( L_F = L - .5p + .25 \) it would be over-contributing unnecessarily.

\[
EU_F(L_F < L - .5p + .25) = -L_F \leq EU_F(L_F = L - .5p + .25) = .25 - L_F
\]

(No incentive for the firm to deviate by contributing less.)

\[
EU_F(L_F > L = L - .5p + .25) = .25 - L_F < EU_F(L_F = L - .5p + .25) = .25 - L_F
\]

(No incentive for the firm to deviate by contributing more)

**Middle Cost Parameter Space:**

\[
\max\{\hat{\theta}_W(\tau_G(0)) + E_G + \hat{\theta}_W(\tau_F(0)), .5p + .25\} \leq L \leq \hat{\theta}_S(\tau_G(0)) + E_G + \hat{\theta}_S(\tau_F(0))
\]
1. Firm’s Strategy:

\[ \sigma_F(m = s) : L_F = \max\{\hat{\theta}_W(\tau_F(0)), L - \hat{\theta}_S(\tau_G(0))\} \]

\[ \sigma_F(m = w) : L_F = 0 \]

The specification of \( L_F \) for the strong case is the threshold requirement that the firm contributes at least as much as it could expect to gain from pursuing a weak case \((\hat{\theta}_W(\tau_F(0)))\), and thus the firm does not have an incentive to bluff with weak cases. The firm may contribute more than \(\hat{\theta}_W(\tau_F(0))\) when the total cost of initiating the case is higher, in which case the firm contributes \(L - \hat{\theta}_S(\tau_G(0))\).

2. Government’s Beliefs: The government’s beliefs about the case are that a case is strong if the firm contributes at least \(\hat{\theta}_W(\tau_F(0))\) and otherwise the government believes the case is weak.

\[ Pr(\theta_S|L_F \geq \hat{\theta}_W(\tau_F(0))) = 1, \]

\[ Pr(\theta_S|L_F < \hat{\theta}_W(\tau_F(0))) = 0 \]

3. Government Strategy: First we consider the government’s strategy when the firm has contributed enough to meet the litigation threshold, \(L_F \geq \hat{\theta}_W(\tau_F(0))\).

If \(L_F \geq \hat{\theta}_W(\tau_F(0))\), then \(Pr(\theta_S|L_F \geq \hat{\theta}_W(\tau_F(0))) = 1\)

\[ EU_G(I_G) = \hat{\theta}_S(\tau_G(0)) - L_G + E_G \]

\[ EU_G(\neg I_G) = \tau_G(1) \]

The government will initiate if the expected utility of initiating is greater than or equal to not initiating.
\[ \Rightarrow I_G \text{ if: } \hat{\theta}_S(\tau_G(0)) - L_G + E_G \geq \tau_G(1) \]

Substitute and rearrange using: \( \tau_G(0) = 1, \tau_G(1) = 0, \hat{\theta}_S = .75, \) and \( L_G = L - L_F \)

\[ \Rightarrow I_G \text{ if: } L_F \geq L - \hat{\theta}_S - E_G \]
\[ \Rightarrow I_G \text{ if: } .75 \geq L - L_F - E_G \]
\[ \Rightarrow I_G \text{ if: } .75 \geq L_G - E_G, \text{ otherwise } \neg I_G \]

Next we consider the government’s strategy when the firm has contributed, but not enough to meet the litigation threshold, \( L_F \leq \hat{\theta}_W(\tau_F(0)) \).

If \( L_F < \hat{\theta}_W(\tau_G(0)) \), then \( Pr(\theta_S | L_F < \hat{\theta}_W(\tau_F(0))) = 0 \)

\[ EU_G(I_G) = \hat{\theta}_W(\tau_G(0)) - L_G + E_G \]
\[ EU_G(\neg I_G) = \tau_G(1) \]
\[ \Rightarrow I_G \text{ if: } \hat{\theta}_W(\tau_G(0)) - L_G + E_G \geq \tau_G(1) \]

Substitute and rearrange using: \( \tau_G(0) = 1, \tau_G(1) = 0, \theta_W = .25, \) and \( L_G = L - L_F \)

\[ \Rightarrow I_G \text{ if: } .25 \geq L - L_F - E_G \]
\[ \Rightarrow I_G \text{ if: } .25 \geq L_G - E_G, \text{ otherwise } \neg I_G \]

Next we consider the government’s strategy when the firm has not contributed (\( L_F = 0 \)).
If $L_F = 0$, then $Pr(\theta_S|L_F < \hat{\theta}_W(\tau_F(0))) = 0$

$$EU_G(I_G) = \hat{\theta}_W(\tau_G(0)) - L_G + E_G$$

$$EU_G(\neg I_G) = \tau_G(1)$$

$\Rightarrow I_G$ if: $\hat{\theta}_W(\tau_G(0)) - L_G + E_G \geq \tau_G(1)$

Substitute and rearrange using: $\tau_G(0) = 1$, $\tau_G(1) = 0$, and $\hat{\theta}_W = .25$

$\Rightarrow I_G$ if: $.25 \geq L_G - E_G$, otherwise $\neg I_G$

4. Sequential Rationality

We now consider whether the beliefs and strategies of the actors are sequentially rational.

If the case is strong and the litigation cost is in the middle range, then the firm will contribute and the government will initiate in equilibrium:

$$EU_F = .75 - L_F, \text{ if } L_F \geq L - E_G - .75$$

and $L_F = \max\{\hat{\theta}_W(\tau_F(0)), L - \hat{\theta}_S(\tau_G(0))\}$

Payoff from deviation:

If the case is strong and the firm deviated and played $L_F', < \max\{\hat{\theta}_W(\tau_F(0)), L - \hat{\theta}_S(\tau_G(0))\}$,

$L - \hat{\theta}_S(\tau_G(0))\}$,

then the government would play $\neg I_G \Rightarrow EU_F = -L_F,$

$\Rightarrow -L_F \leq .75 - L_F \Rightarrow \text{Firm} \neq \text{deviate.}$
If the firm deviated and chose to contribute less than the contribution threshold (including $L_F = 0$), then the government would not initiate and the firm’s expected utility would be $-L_F'$, which is weakly less than the expected utility of initiating ($0 \leq .75 - L_F$), so the firm will not deviate.

Furthermore, the firm would never pay more than the contribution threshold, since any additional expenditure cannot influence the government’s decision to initiate (since the government already chooses to initiate once the threshold is met), and thus any additional contribution only reduces the firm’s expected payoff.

Next we consider whether there is an incentive to deviate from the equilibrium if the case is weak in the middle cost range.

If the case is weak and the litigation cost is in the middle range, then the firm will play $L_F = 0$ and the government will play $\neg I_G$:

$$L_F = 0 \Rightarrow \neg I_G \Rightarrow E U_F = 0$$

Payoff from deviation:

If the case is weak and the firm deviated and played $L_{F'} > 0$ and

$$L_{F'} < \hat{\theta}_W (\tau_F (0)),$$ then the government believes the case is weak and plays $\neg I_G$.

$$\Rightarrow E U_F = -L_{F'},$$

$$-L_{F'} < 0 \Rightarrow \text{Firm} \neq \text{deviate}.$$
old, the government would not initiate the case, and the firm’s expected utility would be $-L_{Fr}$, which is less than 0, so the firm will not deviate.

If the case is weak and the firm deviated and played $L_{Fr} \geq \hat{\theta}_W(\tau_F(0))$,
then the government will believe the case is strong and play $I_G$.

$\Rightarrow EU_F = \hat{\theta}_W(\tau_F(0)) - L_{Fr}$
$\hat{\theta}_W(\tau_F(0)) - L_{Fr} \leq 0 \Rightarrow \text{Firm } \neq \text{deviate.}$

If the firm deviated and contributed at least the contribution threshold, the government would then initiate the case, and the firm’s expected utility would be $\hat{\theta}_W(\tau_F(0)) - L_{Fr}$, which is no better than the payoff for not contributing (0), so the firm will not deviate. This demonstrates that the firm does not have an incentive to bluff and try to convince the government to initiate weak cases.

**High Cost Parameter Space:** $L > \hat{\theta}_S(\tau_G(0)) + E_G + \hat{\theta}_S(\tau_F(0))$

1. Firm’s Strategy:

$\sigma_F(m = s) = \sigma_F(m = w) : L_F = 0$

2. Government’s Beliefs:

$Pr(\theta_S) = p$

The government does not update its prior, since the firm is pooling and thus no information is conveyed.

3. Government’s Strategy:

$\sigma_G = \neg I_G$
$EU_G(\neg I_G) = (\tau_G(1)) = 0$

4. Sequential Rationality

We now consider whether the beliefs and strategies of the actors are sequentially rational.

If the case is strong and the litigation cost is in the high cost range, then the firm will not contribute and the government will not initiate in equilibrium:

$L_F = 0 \Rightarrow \neg I_G \Rightarrow EU_F = (\tau_F(1)) = 0$

Payoff from deviation:

If the case is strong and the firm deviated and played $L_F > 0$, let us consider the maximum amount the firm would ever contribute, $L_F = \tilde{\theta}_S(\tau_F(0))$

(the most the firm would ever contribute, since it is the most it could ever expect to win from the case). Since we are off the path let’s assume the most favorable beliefs of the government for the case, and thus the government believes the case is strong.

$EU_G(I_G) = \tilde{\theta}_S(\tau_G(0)) - L_G$, where $L_G = L - L_F$

$EU_G(\neg I_G) = \tau_G(1) = 0$

Because it is a high cost case, $L_G > \tilde{\theta}_S(\tau_G(0))$

$\Rightarrow \tilde{\theta}_S(\tau_G(0)) - L_G < 0 \Rightarrow$ Firm $\neq$ deviate.
Given that the firm contribution, even when it is at the maximum where the firm could
break-even, cannot convince the government to initiate the case in the high cost range,
there is no incentive for the firm to deviate and contribute in any amount if the case
is in the high cost range.

**A.3 Uniqueness of Equilibrium**

To show the uniqueness of the previously discussed equilibrium, the following pages con-
jecture alternative pooling, mixed, and semi-pooling equilibria and off-the-path beliefs, and
demonstrate that none are sequentially rational.

### A.3.1 Mixed Possibility

Here we consider the possibility that the firm adopts some form of mixed strategy. The
proof shows that the firm has a profitable deviation to a pure strategy, and thus the firm
will not mix.

1. \( \sigma_F(m = s) : Pr(L_F > 0) = y \), \( \sigma_F(m = w) : Pr(L_F > 0) = q \)

This strategy profile is kept as generic as possible, only specifying that the firm makes
some positive contribution with probability \( y \) or \( q \), depending on the message it receives.

2. Government’s Beliefs (using Bayes’ rule):

\[
Pr(\theta_S|L_F > 0) = \frac{yp}{yp + q - pq} = K
\]

\[
Pr(\theta_S|L_F = 0) = \frac{p - yp}{p - yp + 1 - p - q + pq} = J
\]
3. Government Strategy:

If $L_F > 0$, $Pr(\theta_S) = K$

$$EU_G(I_G) = K\hat{\theta}_s(\tau_G(0)) + (1 - K)\hat{\theta}_w(\tau_G(0)) - L_G$$

Substitute and rearrange using: $\hat{\theta}_s = .75$, $\hat{\theta}_w = .25$, $\tau_G(0) = 1$

$$EU_G(I_G) = .75K + .25(1 - K) - L_G$$

$$= .5K + .25 - L_G$$

$$= .5K + .25 - L + L_F$$

$$EU_G(\neg I_G) = 0$$

For the government to be indifferent when the firm contributes, it must be that the expected utilities of initiating and not initiating the dispute are equal, which requires:

**1st Indifference Condition:** $0 = .5K + .25 - L + L_F$

$$\Rightarrow L - L_F - .25 = .5K$$

If $L_F = 0$, $Pr(\theta_S) = J$

$$EU_G(I_G) = .75J + .25(1 - J) - L_G$$

$$= .5J + .25 - L_G$$

$$= .5J + .25 - L$$

$$EU_G(\neg I_G) = 0$$

For the government to be indifferent when the firm does not contribute, it must be that the expected utilities of initiating and not initiating the dispute are equal, which requires:
2nd Indifference Condition: \( 0 = .5J + .25 - L \Rightarrow L - .25 = .5J \)

4. Sequential Rationality:

To show that the mixing strategy is not sequentially rational, the following demonstrates that the firm has a dominant pure strategy when the case is strong. (This is followed by showing that the firm also has a dominant pure strategy when the case is weak.)

If \( m = s \), and the firm contributes \( L_F > 0 \) then:

1. \( EU_F = .75 - L_F \), when \( L - L_F - .25 \leq .5K \) (Government initiates)
2. \( EU_F = -L_F \), when \( L - L_F - .25 > .5K \) (Government does not initiate)

If \( m = s \), and the firm does not contribute, \( L_F = 0 \), then:

3. \( EU_F = .75 \), when \( L - .25 \leq .5J \) (Government initiates)
4. \( EU_F = 0 \), when \( L - .25 > .5J \) (Government does not initiate)

The 1st and 2nd indifference conditions correspond to the threshold conditions in lines 1 and 3 of the sequential rationality proof. Thus, when the indifference conditions are met, the firm is faced with a choice between a payoff of \( .75 - L_F \) (Line 1), or a payoff of \( .75 \) (Line 3). The firm is strictly better off contributing nothing and receiving \( .75 \) and thus the firm is unwilling to mix.

If \( m = w \), and the firm contributes \( L_F > 0 \) then:

5. \( EU_F = .25 - L_F \), when \( L - L_F - .25 \leq .5K \) (Government initiates)
6. \( EU_F = -L_F \), when \( L - L_F - .25 > .5K \) (Government does not initiate)

If \( m = w \), and the firm does not contribute, \( L_F = 0 \), then:
7. $EU_F = .25$, when $L - .25 \leq .5J$ (Government initiates)

8. $EU_F = 0$, when $L - .25 > .5J$ (Government does not initiate)

Similar to the strong case, the 1st and 2nd indifference conditions correspond to the threshold conditions in lines 5 and 7 of the sequential rationality proof. Thus, when the indifference conditions are met, the firm is faced with a choice between a payoff of $.25 - L_F$ (Line 5), or a payoff of $.25$ (Line 7). The firm is strictly better off contributing nothing and receiving $.25$ and thus the firm is unwilling to mix.

Since the firm has a dominant pure strategy when the indifference conditions are satisfied, a mixing equilibrium is not sequentially rational.
A.3.2 Alternative Semi-Pooling Possibilities

It is also possible that the firm could use a semi-pooling strategy, where it mixed strategies only when a case is strong or weak (but not both). Such semi-pooling equilibria would be special cases of the potential mixed equilibria, where $y$ or $q \in \{0, 1\}$. Given that these are a special subset of the previously analyzed mixed equilibria, and the mixed equilibria is not sequentially rational for any values of $y$ and $q$, it follows that the semi-pooling equilibria are not sequentially rational and thus are not sustainable perfect bayesian equilibria.

A.3.3 Alternative Separating Possibility

In this potential alternative equilibrium, we conjecture that the firm could choose to contribute $L_F > 0$ for weak cases, but not for strong cases. The equilibrium does not hold, because the firm deviates when the case is weak.

1. Firms Strategy:

   $\sigma_F(m = s) : L_F = 0$

   $\sigma_F(m = w) : L_F > 0$

2. Government’s Beliefs:

   $Pr(\theta_S|L_F = 0) = 1$

   $Pr(\theta_S|L_F > 0) = 0$

The government updates their beliefs, and given the fully separating strategy of the firm, accurately deduces the strength of the case.
3. Government’s Strategy:

If $L_F > 0$, then:

$$\sigma_G = I_G, \text{ when } .25 - L_G \geq 0; \neg I_G \text{ otherwise.}$$

$$EU_G(I_G|L_F > 0) = \hat{\theta}_W(\tau_G(0)) - L_G = .25 - L_G$$

$$EU_G(\neg I_G|L_F > 0) = 0$$

Government plays $I_G$ if $.25 - L_G \geq 0$; $\neg I_G$ otherwise.

If $L_F = 0$, then:

$$\sigma_G = I_G, \text{ when } .75 \geq L ; \neg I_G \text{ otherwise.}$$

$$EU_G(\neg I_G|L_F = 0) = 0$$

$$EU_G(I_G|L_F = 0) = \hat{\theta}_S(\tau_G(0)) - L_G = .75 - L$$

Government plays $I_G$ if $.75 \geq L$; $\neg I_G$ otherwise.

4. Sequential Rationality:

We now consider whether the beliefs and strategies of the actors are sequentially rational. If the case is weak, the equilibrium conjecture says that the firm will contribute:

Consider the cases when the total litigation cost is less than the expected payoff of a strong case: $L \leq \hat{\theta}_S(\tau_j(0)) \Rightarrow L \leq 0.75$
The only incentive for the firm to contribute is to encourage the government to initiate
the case, which requires $L_F \geq L - .25$. The best payoff the firm can receive in this
situation is:

$$EU_F(L_F > 0) = \hat{\theta}_W(\tau_F(0)) - L_F = .25 - L_F$$

If the firm deviates and chooses not to contribute ($L_F = 0$), then the government
infers the case is strong and chooses to initiate the case (given $L \leq .75$), which results
in the following expected payoff to the firm:

$$EU_F(L_F = 0) = \hat{\theta}_W(\tau_F(0)) = .25$$

$$EU_F(L_F = 0) = .25 < .25 - L_F = EU_F(L_F > 0)$$

The expected payoff to the firm for deviating and not contributing is strictly greater
than the payoff for contributing, and thus the firm will deviate, so the equilibrium is
not sequentially rational.

**A.3.4 Pooling Possibility 1a**

In this potential alternative equilibrium, we conjecture that the firm contributes for both
weak and strong cases. The equilibrium does not hold, because firms would deviate when the
case is strong, such that the government would initiate the case without the firm contributing.
We consider this equilibrium under multiple belief structures. In each set of beliefs the government believes the case is strong with probability $p$ when the firm contributes (on the path). We then look at two extremes for off the path beliefs, where the government first believes the case is strong when the firm does not contribute, followed by the belief that the case is weak when the firm does not contribute.

1. Firm’s Strategy:

$$\sigma_F(m = s) = \sigma_F(m = w) : L_F = L - p \hat{\theta}_S(\tau_G(0)) + (1 - p) \hat{\theta}_W(\tau_G(0)),$$

when $L > p \hat{\theta}_S(\tau_G(0)) + (1 - p) \hat{\theta}_W(\tau_G(0)); \ L_F = \epsilon \to 0$ otherwise.

This states that the firm contributes just enough to make the government indifferent between initiating the case and not initiating, or if the total litigation cost is less than the expected payoff to the government of initiating, the firm contributes some arbitrarily small amount.

2. Government’s Beliefs:

$$Pr(\theta_S|L_F > 0) = p$$

$$Pr(\theta_S|L_F = 0) = 1$$

In this belief structure, the government maintains their prior when the firm contributes since the firm has a pooling strategy and contributes on the path, but updates their belief that the case is strong if the firm does not contribute off the path (see the following proof for alternative beliefs).

3. Government Strategy:

If $L_F > 0$, then:
\[ \sigma_G = I_G \text{ when } L_G - .25 \leq .5p; \neg I_G \text{ otherwise.} \]

If \( L_F > 0 \), \( Pr(\theta_S) = p \)

\[ EU_G(I_G) = p(\hat{\theta}_S(\tau_G(0)) - L_G) + (1 - p)(\hat{\theta}_W(\tau_G(0)) - L_G) = .5p + .25 - L_G \]

\[ EU_G(\neg I_G) = 0 \]

\( I_G \) if \( L_G - .25 \leq .5p; \neg I_G \) otherwise.

If \( L_F = 0 \), then:

\[ \sigma_G = I_G \text{ when } .75 \geq L; \neg I_G \text{ otherwise.} \]

If \( L_F = 0 \), \( Pr(\theta_S) = 1 \)

\[ EU_G(I_G) = \hat{\theta}_S(\tau_G(0)) - L_G = .75 - L \]

\( I_G \) if \( .75 \geq L; \neg I_G \) otherwise.

4. Sequential Rationality:

We now consider whether the beliefs and strategies are sequentially rational.

The equilibrium conjectures that the firm will always contribute \( L_F > 0 \).

Consider the situation when the case is strong and \( L \leq \hat{\theta}_S(\tau_j(0)) \Rightarrow L \leq 0.75 \).

If the firm contributes \( L_F > 0 \) such that \( L_G - .25 \leq .5p \) then the government will initiate the case. This results in the following expected payoff to the firm:
\[ EU_F(L_F > 0) = \hat{\theta}_S(\tau_F(0)) - L_F = .75 - L_F \]

Payoff from deviation:
\[ EU_F(L_F = 0) = \hat{\theta}_S(\tau_F(0)) = .75 \]
\[ EU_F(L_F = 0) = .75 > .75 - L_F = EU_F(L_F > 0) \]

The expected payoff to the firm for deviating and not contributing is strictly greater than the payoff for contributing, and thus the firm will deviate, so the equilibrium is not sequentially rational.

**A.3.5 Pooling Possibility 1b**

We now examine the same strategy profile for the firm, but under an alternative belief structure.

1. Firm’s Strategy:
   \[ \sigma_F(m = s) = \sigma_F(m = w) : L_F = L - p \hat{\theta}_S(\tau_G(0)) + (1 - p) \hat{\theta}_W(\tau_G(0)), \]
   when \( L > p \hat{\theta}_S(\tau_G(0)) + (1 - p) \hat{\theta}_W(\tau_G(0)) \); \( L_F = \epsilon \rightarrow 0 \) otherwise.

   This states that the firm contributes just enough to make the government indifferent between initiating the case and not initiating, or if the total litigation cost is less than the expected payoff to the government of initiating, the firm contributes some arbitrarily small amount.

2. Government’s Beliefs:
   \[ Pr(\theta_S|L_F > 0) = p \]
\[ Pr(\theta_S | L_F = 0) = 0 \]

In this belief structure, the government maintains their prior when the firm contributes since the firm has a pooling strategy and contributes on the path, but updates their belief that the case is weak if the firm does not contribute off the path.

3. Government Strategy:

If \( L_F > 0 \), then:

\[ \sigma_G = I_G \text{ when } L_G - .25 \leq .5p; \neg I_G \text{ otherwise.} \]

If \( L_F > 0 \), \( Pr(\theta_S) = p \)

\[ EU_G(I_G) = p(\hat{\theta}_S(\tau_G(0)) - L_G) + (1 - p)(\hat{\theta}_W(\tau_G(0)) - L_G) = .5p + .25 - L_G \]

\[ EU_G(\neg I_G) = 0 \]

\( I_G \) if \( L_G - .25 \leq .5p; \neg I_G \) otherwise.

If \( L_F = 0 \), then:

\[ \sigma_G = I_G \text{ when } .25 \geq L; \neg I_G \text{ otherwise.} \]

If \( L_F = 0 \), \( Pr(\theta_S) = 0 \)

\[ EU_G(I_G) = \hat{\theta}_W(\tau_G(0)) - L_G = .25 - L_G = .25 - L \]

\[ EU_G(\neg I_G) = 0 \]

\( I_G \) if \( .25 \geq L \); \( \neg I_G \) otherwise.
4. Sequential Rationality:

We now consider whether the beliefs and strategies are sequentially rational.

Consider the situation when the case is weak and:

\[
\hat{\theta}_W(\tau_F(0)) < L - p \hat{\theta}_S(\tau_G(0)) + (1 - p) \hat{\theta}_W(\tau_G(0)), \quad \text{and} \quad L > p \hat{\theta}_S(\tau_G(0)) + (1 - p) \hat{\theta}_W(\tau_G(0))
\]

The firm’s strategy is to contribute the minimum amount necessary to make the government indifferent between initiating and not, which is:

\[
L_F = L - p \hat{\theta}_S(\tau_G(0)) + (1 - p) \hat{\theta}_W(\tau_G(0)) > .25.
\]

Having contributed enough, the government will then initiate the case. This results in the following expected payoff to the firm:

\[
EU_F(L_F) = L - p \hat{\theta}_S(\tau_G(0)) + (1 - p) \hat{\theta}_W(\tau_G(0)) - L_F = \hat{\theta}_W(\tau_F(0)) - L_F = .25 - L_F < 0
\]

Because the firm contribution is greater than the firms expected payoff for a weak case, the expected payoff to the firm is strictly negative.

Payoff from deviation:
If the firm chooses to contribute nothing, the government will not initiate the case and the expected payoff to the firm is:

\[ EU_F(L_F = 0) = 0 \]

The expected payoff to the firm for deviating and not contributing is strictly greater than the payoff for contributing, and thus the conjectured equilibrium is not sequentially rational.

A.3.6 Pooling Possibility 2a

In this potential alternative equilibrium, we conjecture that the firm does not contribute \((L_F = 0)\) for both weak and strong cases. The equilibrium does not hold, because the firm would deviate when the case is strong and a contribution up to the firm’s expected payoff of a strong case is sufficient to convince the government to initiate the case.

We consider this equilibrium under multiple belief structures. In each set of beliefs the government believes the case is strong with probability \(p\) when the firm does not contribute (on the path). We then look at two extremes for off the path beliefs, where the government first believes the case is strong when the firm does not contribute, followed by the belief that the case is weak when the firm does not contribute.

1. Firm’s Strategy:

\[ \sigma_F(m = s) = \sigma_F(m = w) : L_F = 0 \]

2. Government’s Beliefs:

\[ Pr(\theta_S|L_F = 0) = p \]
\[ Pr(\theta_S|L_F > 0) = 1 \]

In this belief structure, the government maintains their prior when the firm does not contribute since the firm has a pooling strategy and does not contribute on the path, but updates their belief that the case is strong if the firm does contribute off the path.

3. Government Strategy:

If \( L_F = 0 \), then:
\[
\sigma_G = I_G \text{ when } L \leq .5p + \tilde{\theta}_W(\tau_G(0)); \neg I_G \text{ otherwise.}
\]

If \( L_F = 0 \), \( Pr(\theta_S) = p \)
\[
EU_G(I_G) = p \left( \tilde{\theta}_S(\tau_G(0)) - L_G \right) + (1 - p)\left( \tilde{\theta}_W(\tau_G(0)) - L_G \right) = .5p + .25 - L_G
\]
\[
= .5p + .25 - L
\]
\[
EU_G(\neg I_G) = 0
\]
\[
I_G \text{ if } L \leq .5p + .25
\]

If \( L_F > 0 \), then:
\[
\sigma_G = I_G \text{ when } L - \tilde{\theta}_S(\tau_G(0)) \leq L_F; \neg I_G \text{ otherwise.}
\]

If \( L_F > 0 \), \( Pr(\theta_S) = 1 \)
\[
EU_G(I_G) = \tilde{\theta}_S(\tau_G(0)) - L_G = .75 - (L - L_F) = .75 - L + L_F
\]
\[
EU_G(\neg I_G) = 0
\]
\[
I_G \text{ if } L_F \geq L - .75, \neg I_G \text{ otherwise.}
\]
4. Sequential Rationality:

We now consider whether the beliefs and strategies are sequentially rational.

Consider the situation when the case is weak and:

$$0.5p + \theta_W(\tau_j(0)) < L < \theta_W(\tau_j(0)) + \hat{\theta}_S(\tau_j(0)) = 1$$

The firm’s strategy is to not contribute ($L_F = 0$). Because $L > 0.5p + \theta_W(\tau_j(0))$, the government does not initiate the case. This results in the following expected payoff to the firm:

$$EU_F(L_F = 0) = 0$$

Payoff from deviation:

If the firm chooses to contribute, it will contribute the minimum amount to make the government indifferent between initiating and not initiating, which is:

$$L_F = L - \hat{\theta}_S(\tau_G(0)) = L - .75$$

Because the government believes the case is strong when the firm contributes, the government initiates the case once the firm contributes, which results in the following expected payoff to the firm:
\[ EU_F(L_F = L - 0.75) = \hat{\theta}_W(\tau_F(0)) - L_F = 0.25 - L_F \]

Given that \( L < \hat{\theta}_W(\tau_j(0)) + \hat{\theta}_S(\tau_j(0)) \Rightarrow L < 1 \)

\[ \Rightarrow L_F = L - 0.75 < 0.25 \Rightarrow L_F < 0.25 \]

\[ \Rightarrow EU_F(L_F = L - 0.75) = 0.25 - L_F > 0 = EU_F(L_F = 0) \]

The expected payoff for the firm deviating and contributing just enough to convince the government to initiate the case is strictly greater than the payoff for not contributing, and thus the conjectured equilibrium is not sequentially rational.

**A.3.7 Pooling Possibility 2b**

We now examine the same strategy profile for the firm, but under an alternative belief structure.

1. Firm’s Strategy:
   \[ \sigma_F(m = s) = \sigma_F(m = w) : L_F = 0 \]

2. Government’s Beliefs:
   \[ Pr(\theta_S|L_F = 0) = p \]
   \[ Pr(\theta_S|L_F > 0) = 0 \]

   In this belief structure, the government maintains their prior when the firm does not contribute since the firm has a pooling strategy and does not contribute on the path,
but updates their belief that the case is weak if the firm does contribute off the path.

3. Government Strategy:

If $L_F = 0$, then:

$\sigma_G = I_G$ when $L \leq 0.5p + \hat{\theta}_W(\tau_G(0))$; $\neg I_G$ otherwise.

If $L_F = 0$, $Pr(\theta_S) = p$

$EU_G(I_G) = p (\hat{\theta}_S(\tau_G(0)) - L_G) + (1 - p)(\hat{\theta}_W(\tau_G(0)) - L_G) = 0.5p + 0.25 - L_G$

$EU_G(\neg I_G) = 0$

$I_G$ if $L \leq 0.5p + 0.25$

If $L_F > 0$, then:

$\sigma_G = I_G$ when $L - \hat{\theta}_W(\tau_G(0)) \leq L_F$; $\neg I_G$ otherwise.

If $L_F > 0$, $Pr(\theta_S) = 0$

$EU_G(I_G) = \hat{\theta}_W(\tau_G(0)) - L_G = 0.25 - (L - L_F) = 0.25 - L + L_F$

$EU_G(\neg I_G) = 0$

$I_G$ if $L_F \geq L - 0.25$, $\neg I_G$ otherwise.

4. Sequential Rationality:
We now consider whether the beliefs and strategies are sequentially rational.

Consider the situation when the case is weak and:

\[ .5p + \hat{\theta}_W(\tau_j(0)) < L < \hat{\theta}_W(\tau_j(0)) + \hat{\theta}_S(\tau_j(0)) = 1 \]

The firm’s strategy is to not contribute \((L_F = 0)\). Because \(L > .5p + \hat{\theta}_W(\tau_j(0))\), the government does not initiate the case. This results in the following expected payoff to the firm:

\[
EU_F(L_F = 0) = 0
\]

Payoff from deviation:

If the firm chooses to contribute, it will contribute the minimum amount to make the government indifferent between initiating and not initiating, which is:

\[
L_F = L - \hat{\theta}_W(\tau_G(0)) = L - .25 \Rightarrow L_G = \hat{\theta}_W(\tau_G(0))
\]

Because the government believes the case is weak when the firm contributes, the government only initiates the case once the firm contributes enough, such that \(L_G \leq \hat{\theta}_W(\tau_G(0))\), which results in the following expected payoff to the firm:

\[
EU_F(L_F = L - .25) = \hat{\theta}_S(\tau_F(0)) - L_F = .75 - L_F
\]
Given that $L < \hat{\theta}_W(\tau_j(0)) + \hat{S}(\tau_j(0)) \Rightarrow L < 1$

$\Rightarrow L_F = L - .25 < .75 \Rightarrow L_F < .75$

$\Rightarrow EU_F(L_F = L - .25) = .75 - L_F > 0 = EU_F(L_F = 0)$

The expected payoff for the firm deviating and contributing just enough to convince the government to initiate the case is strictly greater than the payoff for not contributing, and thus the conjectured equilibrium is not sequentially rational.

A.4 Multiple Firms with Incomplete Information

This section discusses the role of private firm contributions when there are heterogenous firms within an industry and uncertainty among the firms about how much they each value initiating the WTO dispute. Rather than focusing on the government’s uncertainty about the strength of the case, this extension holds the strength of the case constant and examines how uncertainty over heterogeneous valuations by firms within an industry affect the likelihood that firms contribute a sufficient amount for the case to be brought.

Building from the previously discussed model, I examine the set of cases where the litigation cost is sufficiently high such that the government will not initiate the case on its own ($L > .5p + .25$). It is helpful to consider the government’s decision to initiate the dispute as a provision of a public good, where each firm in the industry values initiating the case at $V_i = \tau_i(0) - \tau_i(1)$\(^1\) The values for individual firms are independently drawn from a

\(^1\)In the two-player game, the value would be multiplied by the probability of winning the case; however, since this extension holds the strength of the case constant, this component is dropped from the analysis since it does not affect the comparative statics.
continuous distribution \( F \), and each firm only knows its own value, although each knows the distribution from which the values were drawn.\footnote{In this way, each firm has private information about their value, although they all share a common distribution within the industry, which provides a level of commonality for all firms within a given industry.}

In the two player game, it was shown that for given parameters, the firm could contribute \( L_F^* \) which was the necessary threshold for the government to initiate the case. In this extension, \( L_F^* \) is the cost of the “public good,” or the necessary contribution threshold that the firms must reach for the government to bring the case. I allow \( L_F \) to be the sum of litigation contributions \((L_i)\) from all firms within the industry. Because firms’ litigation contributions involve sinking costs into the litigation process through fact finding and preparation of materials, I consider firms’ contributions to be non-refundable in the model. This means that if firms contribute and fail to reach the necessary threshold \( (L_F^*) \), the costs are sunk.

Given this setup, which is based on a set of realistic assumptions drawing upon how the WTO litigation process functions, this game is best described as a contribution game with uncertainty and heterogenous preferences. The game is formalized through the existence of \( N \geq 2 \) firms, where each firm \( i, i, ..., N \), only knows his own value \((V_i)\) for initiating the case, which will be brought if the the firms contribute a combined \( L_F^* \). This type of game has been analyzed in the generic form (for the provision of any discrete public good) by Menezes, Monteiro, and Temimi (2001) in “Private Provision of Discrete Public Goods with Incomplete Information.”

Since the game has been thoroughly analyzed elsewhere, I draw from the earlier insights and discuss the key implications for dispute initiation in the WTO. To illustrate the connection between between the specifics of the game here and the work of Menezes, Monteiro, and Temimi (2001), the following lines specify how the games are linked.

1. In each game there are \( N \geq 2 \), where \( i, i, ..., N \), knows its own value \((V_i)\), but only knows the distribution of others’ values \((F)\).
2. The cost of providing the public good for Menezes, Monteiro, and Temimi (2001, 496) is $c$, which is equivalent to $L_r^*F$. This holds, given that for any set of constant parameters there exists an $L_r^*$ such that the government will initiate the case.

3. In each game, the individual players make a simultaneous decisions to contribute, where the contribution is any amount greater than or equal to zero ($L_i \geq 0$).

4. The “public good” is dichotomous, as shown by the government’s decisions to either initiate ($I_G$) or not initiate ($\neg I_G$) the case.

In this set up, Menezes, Monteiro, and Temimi (2001) prove the following theorem.

**Theorem 1.** Suppose $F: [0, \infty) \to \mathbb{R}$ is a continuous distribution. Suppose there are $N \geq 2$ players for a project with cost $c > 0$ and that $F(c) < 1$. Then there exists an $\alpha > 0$, where $\alpha$ solves $\alpha F(\alpha)^{N-1} = c$ such that

$$
  b(v) = \begin{cases} 
  0, & \text{if } v \leq \alpha \\
  c, & \text{if } v > \alpha,
  \end{cases}
$$

is an equilibrium strategy for the contribution game.

Theorem 1 implies that when the cost of the provision is not prohibitively high as to prevent a single player from providing the good, there always exists an equilibrium where a player with a sufficiently large valuation provides the good himself. (emphasis in original)

From this, we can consider the situation when there are multiple industries, each with multiple firms. All else equal, in expectation the industry that has the firm with the highest
valuation (and ability to contribute) will be the most likely to have at least one firm where \( v > \alpha \), and is thus the most likely industry in which a firm would contribute and a dispute would be initiated. Based on this, in the empirical section I examine how the size of dominant firms within industries affects the likelihood of WTO dispute initiation.

Menezes, Monteiro, and Temimi (2001, 503) also show that, when no single firm can afford to pay the cost of providing the public good, and the cost of the good is high enough (“slightly above the aggregate mean of the valuations”) then the unique equilibrium of the game is to contribute nothing.

Theorem 2. If the public project cost is higher than \( C_N \), then the unique equilibrium of the contribution game is the strong free riding equilibrium, i.e.

\((b_1(\cdot), ..., b_N(\cdot))\), where \(b_i(v) \equiv 0\).

This theorem implies that as the average value for firms within an industry declines, it is increasingly likely that they contribute nothing and the dispute will not be initiated. Conversely, as the average value for firms within an industry increases, it is increasingly likely that they contribute and the dispute will be initiated. Based on this, in the empirical section I examine how the average size of firms within industries affects the likelihood of WTO dispute initiation.

A.5 Robustness Checks

A.5.1 Sample Variation in Empirical Models

Due to data limitations, the number of observations varies in the empirical analysis of the paper. In the following table, the models from Table 1 are replicated, but use the same constrained sample across all models. The results show that the main results are not an artifact of the changing samples across models.
Table A.1: Random Effects Logistic Regression of WTO Dispute Complaints

<table>
<thead>
<tr>
<th></th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
<th>Model 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dominant Firm Capacity</td>
<td>0.298**</td>
<td>0.332**</td>
<td>0.302**</td>
<td>0.259*</td>
<td>0.310*</td>
</tr>
<tr>
<td></td>
<td>(0.151)</td>
<td>(0.161)</td>
<td>(0.154)</td>
<td>(0.159)</td>
<td>(0.175)</td>
</tr>
<tr>
<td>Trade Barrier Distortion</td>
<td>2.235***</td>
<td>2.335***</td>
<td>2.257***</td>
<td>2.082***</td>
<td>2.238***</td>
</tr>
<tr>
<td></td>
<td>(0.772)</td>
<td>(0.796)</td>
<td>(0.776)</td>
<td>(0.777)</td>
<td>(0.807)</td>
</tr>
<tr>
<td>Negotiation Progress</td>
<td>−0.977**</td>
<td>−0.952**</td>
<td>−0.919**</td>
<td>−1.035**</td>
<td>−1.010**</td>
</tr>
<tr>
<td></td>
<td>(0.464)</td>
<td>(0.476)</td>
<td>(0.466)</td>
<td>(0.466)</td>
<td>(0.475)</td>
</tr>
<tr>
<td>Trade Barrier Duration</td>
<td>−0.239*</td>
<td>−0.234*</td>
<td>−0.244*</td>
<td>−0.240*</td>
<td>−0.134</td>
</tr>
<tr>
<td></td>
<td>(0.136)</td>
<td>(0.138)</td>
<td>(0.136)</td>
<td>(0.134)</td>
<td>(0.153)</td>
</tr>
<tr>
<td>EU</td>
<td>1.021</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.113)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Japan</td>
<td>0.682</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.277)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mexico</td>
<td>1.130</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.473)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Korea</td>
<td>0.488</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.263)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NonOECD</td>
<td>0.028</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.181)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>US Exports to Trade Partner</td>
<td>0.146</td>
<td></td>
<td></td>
<td>−2.016</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.235)</td>
<td></td>
<td></td>
<td>(1.294)</td>
<td></td>
</tr>
<tr>
<td>Industry Production</td>
<td>−0.055</td>
<td></td>
<td>0.310</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.269)</td>
<td></td>
<td>(0.444)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Political Contributions</td>
<td>−0.165</td>
<td>−0.255</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.216)</td>
<td>(0.328)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Active 301</td>
<td>1.335*</td>
<td>1.468</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.718)</td>
<td>(0.984)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

N: 1407

*p < .1, **p < .05, ***p < .01

This table reports results using the smallest subset of data with results reported in Table 1 of the main paper. Random effect models calculated using xtmelogit with STATA10. Random intercepts calculated for groups at the industry level, defined as the ISIC3 4 digit industry. Canada is the omitted comparison.

A.5.2 Industry Level Fixed Effects

It is possible that certain industries are more or less likely to engage in trade disputes, regardless of dominant firm capacity. To address this concern the following table replicates
the models from Table 1, but uses fixed effects models, with fixed effects for each industry. This allows us to examine how changes in the variables of interest affect dispute initiation within industries. The results are consistent with those reported in the body of the paper, showing that Dominant Firm Capacity is not just capturing other traits of the industry.

Table A.2: Industry Fixed Effects Logistic Regression of WTO Dispute Complaints

<table>
<thead>
<tr>
<th></th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
<th>Model 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dominant Firm Capacity</td>
<td>0.662**</td>
<td>0.776**</td>
<td>0.653**</td>
<td>0.614**</td>
<td>0.889**</td>
</tr>
<tr>
<td></td>
<td>(0.282)</td>
<td>(0.345)</td>
<td>(0.324)</td>
<td>(0.286)</td>
<td>(0.394)</td>
</tr>
<tr>
<td>Trade Barrier Distortion</td>
<td>2.784***</td>
<td>2.886***</td>
<td>2.579***</td>
<td>2.582***</td>
<td>2.722***</td>
</tr>
<tr>
<td></td>
<td>(0.800)</td>
<td>(0.881)</td>
<td>(0.805)</td>
<td>(0.814)</td>
<td>(0.989)</td>
</tr>
<tr>
<td>Negotiation Progress</td>
<td>-1.299***</td>
<td>-1.133**</td>
<td>-1.091**</td>
<td>-1.275***</td>
<td>-1.044**</td>
</tr>
<tr>
<td></td>
<td>(0.463)</td>
<td>(0.478)</td>
<td>(0.486)</td>
<td>(0.472)</td>
<td>(0.505)</td>
</tr>
<tr>
<td>Trade Barrier Duration</td>
<td>-0.225*</td>
<td>-0.208</td>
<td>-0.220*</td>
<td>-0.211</td>
<td>-0.051</td>
</tr>
<tr>
<td></td>
<td>(0.129)</td>
<td>(0.129)</td>
<td>(0.131)</td>
<td>(0.135)</td>
<td>(0.166)</td>
</tr>
<tr>
<td>EU</td>
<td>1.074</td>
<td></td>
<td></td>
<td></td>
<td>2.431</td>
</tr>
<tr>
<td></td>
<td>(1.269)</td>
<td></td>
<td></td>
<td></td>
<td>(1.906)</td>
</tr>
<tr>
<td>Japan</td>
<td>0.420</td>
<td></td>
<td></td>
<td>-0.170</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.438)</td>
<td></td>
<td></td>
<td>(2.342)</td>
<td></td>
</tr>
<tr>
<td>Mexico</td>
<td>1.483</td>
<td></td>
<td></td>
<td>-0.144</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.386)</td>
<td></td>
<td></td>
<td>(2.540)</td>
<td></td>
</tr>
<tr>
<td>Korea</td>
<td>-0.602</td>
<td></td>
<td>-5.179</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.414)</td>
<td></td>
<td>(4.198)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NonOECD</td>
<td>-0.237</td>
<td></td>
<td>-6.439</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.336)</td>
<td></td>
<td>(5.612)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>US Exports to Trade Partner</td>
<td>0.264</td>
<td></td>
<td>-2.653</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.301)</td>
<td></td>
<td>(1.761)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Industry Production</td>
<td>-0.124</td>
<td>0.071</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.023)</td>
<td></td>
<td>(1.731)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Political Contributions</td>
<td>-0.218</td>
<td>0.189</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.870)</td>
<td></td>
<td>(1.123)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Active 301</td>
<td>1.308</td>
<td>1.501</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.909)</td>
<td></td>
<td>(1.432)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| N                        | 1059         | 1059         | 999          | 1056         | 996          |

*p < .1, **p < .05, ***p < .01

Fixed effect models calculated using xtabiv with STATA10. Fixed effects are at the industry level, defined as the ISIC3 2 digit industry. Canada is the omitted comparison.

3The industry is grouped at the ISIC 2 digit level for these models. This is done because 51 groups at the ISIC 4 digit level would have to be dropped due to lack of variation in the dependent variable.
Appendix B

When to Compromise Appendix

B.1 Sample Demographics

Table B.1: Sample Populations and National Population Comparison

<table>
<thead>
<tr>
<th></th>
<th>Qualtrics</th>
<th>Experiment 1</th>
<th>Experiment 2</th>
<th>National Population</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>AGE</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percent Age 18-24</td>
<td>6.3</td>
<td>25.4</td>
<td>19.1</td>
<td>12.9</td>
</tr>
<tr>
<td>Percent Age 25-44</td>
<td>54.2</td>
<td>63.1</td>
<td>63.6</td>
<td>40.5</td>
</tr>
<tr>
<td>Percent Age 45-64</td>
<td>33.9</td>
<td>10.3</td>
<td>15.7</td>
<td>29.6</td>
</tr>
<tr>
<td>Percent Age 65+</td>
<td>5.6</td>
<td>1.2</td>
<td>1.6</td>
<td>16.7</td>
</tr>
<tr>
<td><strong>INCOME</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percent $0-$50,000</td>
<td>40.8</td>
<td>60.0</td>
<td>56.1</td>
<td>50.5</td>
</tr>
<tr>
<td>Percent $50,000-$100,000</td>
<td>36.9</td>
<td>30.7</td>
<td>34.8</td>
<td>29.1</td>
</tr>
<tr>
<td>Percent $100,000-$150,000</td>
<td>14.0</td>
<td>5.8</td>
<td>5.6</td>
<td>12.1</td>
</tr>
<tr>
<td>Percent $150,000-$200,000</td>
<td>5.8</td>
<td>2.2</td>
<td>2.2</td>
<td>4.5</td>
</tr>
<tr>
<td>Percent $200,000+</td>
<td>2.5</td>
<td>1.3</td>
<td>1.5</td>
<td>3.9</td>
</tr>
<tr>
<td><strong>EDUCATION</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Less than High School</td>
<td>1.2</td>
<td>0.5</td>
<td>0.7</td>
<td>14.8</td>
</tr>
<tr>
<td>High School / GED</td>
<td>15.9</td>
<td>12.7</td>
<td>10.9</td>
<td>10.9</td>
</tr>
<tr>
<td>Some College</td>
<td>24.4</td>
<td>31.1</td>
<td>29.4</td>
<td>21.4</td>
</tr>
<tr>
<td>College Degree</td>
<td>46.1</td>
<td>47.5</td>
<td>48.4</td>
<td>25.1</td>
</tr>
<tr>
<td>Masters Degree</td>
<td>9.8</td>
<td>5.7</td>
<td>8.0</td>
<td>7.2</td>
</tr>
<tr>
<td>PhD / JD / MD</td>
<td>2.6</td>
<td>2.5</td>
<td>2.5</td>
<td>3.1</td>
</tr>
</tbody>
</table>

* National age, income, and education percentages are from the 2010 Census.
  Age percentages are only for those age 18 and older. Education percentages are only for those age 25 and older.
### B.2 Comparing mTurk and Qualtrics Results

Table B.2: Comparison of Effects in mTurk Experiment 2 and Qualtrics Samples

<table>
<thead>
<tr>
<th>Stay Out</th>
<th>Engage</th>
<th>Compromise</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mTurk</td>
<td>Qualtrics</td>
</tr>
<tr>
<td>Democrats</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not Engage</td>
<td>1.47***</td>
<td>1.00***</td>
</tr>
<tr>
<td></td>
<td>(7.53)</td>
<td>(3.31)</td>
</tr>
<tr>
<td>Republicans</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not Engage</td>
<td>1.60***</td>
<td>1.06**</td>
</tr>
<tr>
<td></td>
<td>(4.25)</td>
<td>(2.53)</td>
</tr>
</tbody>
</table>

**Note:** ∗p<0.1; ∗∗p<0.05; ∗∗∗p<0.01

Table B.2 shows the difference in average approval scores \((Column \ Score - Row \ Score)\) and t-statistics among Democrats and Republicans based on strategies for the mTurk and Qualtrics samples. In the mTurk sample, territorial and casualty outcomes are constant at no casualties and 20% territory gained by the attacking country. There were no casualty or territory conditions in the Qualtrics sample. The approval score is measured on a scale from -3=Strongly Disapprove to 3=Strongly Approve in both experiments.

Table B.2 illustrates that the effects are remarkably consistent across the mTurk and Qualtrics samples. In all cases the sign of the effect is consistent across samples and both studies find similar statistical significance for the difference in means estimates. Of particular note is that in the national Qualtrics sample, compromise receives even higher relative approval, suggesting that the core effect may be even stronger in a national sample. The Qualtrics sample is not used for the primary analysis, since it does not control for territory and casualty outcomes.
B.3 The Impact of Holding Casualties Constant

In order to isolate the treatment effect of the leader’s strategy, it is necessary that casualties are held constant at zero across all conditions, otherwise respondents’ approval could shift because of different outcomes as opposed to strategies. In the control and treatment conditions that do not result in military engagement (Stay Out, Not Engage, and Compromise) casualties are obviously at zero, and thus no mention of casualties is included. However, in the Engage condition respondents may assume higher costs resulting from war\textsuperscript{1} even if outcomes are not explicitly varied (for a related point, see Sher and McKenzie (2006) and Tomz and Weeks (2013) on “information leakage”). To avoid conflating respondent’s expectations about the human cost of war with approval for the leader’s strategy, I specify that casualties are zero in the engage condition. This inclusion means that the treatment conditions are not lexically equivalent, but instead they are logically equivalent. This technique was also used by Kertzer and Brutger (2014) who also tested the effect of making the conditions lexically equivalent by omitting any mention of casualties in the Engage condition. Kertzer and Brutger found that audiences do indeed impose a higher cost when casualties are not mentioned, which means I am choosing a more conservative measure here, since any bias introduced by specifying casualties at zero would make the Engage treatment relatively more appealing to audiences (meaning that Compromise would be relatively less appealing).

\textsuperscript{1}This would be expected, given that the crisis bargaining model explicitly includes a cost of war that is separate from the division of the contested prize.
B.4 Average Approval Not Conditioned on Party of the President

Figure B.1: Average Approval Score for Each Strategy

![Figure B.1](image)

Figure B.1 shows the average approval score and 90 percent confidence intervals for experiments 1 and 2 for each randomly assigned crisis strategy. Outcomes are held constant with no casualties and 20% territory gained by the attacking country. The approval score is measured on a scale from -3=Strongly Disapprove to 3=Strongly Approve.
### B.5 Full Results for Experiments 1 and 2

Table B.3: Experiment 1: Differences in Approval Score for Strategies and Party

<table>
<thead>
<tr>
<th>Audience of Democrats &amp; President who is Democrat</th>
<th>Difference in Means</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Stay Out</td>
<td>Engage</td>
<td>Not Engage</td>
<td>Compromise</td>
</tr>
<tr>
<td>Stay Out</td>
<td>-0.42</td>
<td>-2.17***</td>
<td>-0.60</td>
</tr>
<tr>
<td>(0.75)</td>
<td>(-0.75)</td>
<td>(-4.94)</td>
<td>(-1.50)</td>
</tr>
<tr>
<td>Engage</td>
<td>-1.75***</td>
<td>-0.18</td>
<td></td>
</tr>
<tr>
<td>(0.33)</td>
<td>(-3.11)</td>
<td>(0.33)</td>
<td>(0.33)</td>
</tr>
<tr>
<td>Not Engage</td>
<td>1.57***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(3.94)</td>
<td>(0.33)</td>
<td>(3.94)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Audience of Democrats &amp; President who is Republican</th>
<th>Difference in Means</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Stay Out</td>
<td>-1.33***</td>
<td>-1.71***</td>
<td>-0.37</td>
</tr>
<tr>
<td>(2.82)</td>
<td>(-2.82)</td>
<td>(-3.65)</td>
<td>(-1.03)</td>
</tr>
<tr>
<td>Engage</td>
<td>-0.38</td>
<td>0.96**</td>
<td></td>
</tr>
<tr>
<td>(0.73)</td>
<td>(-0.73)</td>
<td>(2.23)</td>
<td>(2.23)</td>
</tr>
<tr>
<td>Not Engage</td>
<td>1.34***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(3.15)</td>
<td>(3.15)</td>
<td>(3.15)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Audience of Republicans &amp; President who is Republican</th>
<th>Difference in Means</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Stay Out</td>
<td>0.33</td>
<td>-2.00**</td>
<td>-0.93</td>
</tr>
<tr>
<td>(0.28)</td>
<td>(0.28)</td>
<td>(-2.31)</td>
<td>(-1.22)</td>
</tr>
<tr>
<td>Engage</td>
<td>-2.33*</td>
<td>-1.26</td>
<td></td>
</tr>
<tr>
<td>(0.91)</td>
<td>(-1.91)</td>
<td>(-1.10)</td>
<td>(-1.10)</td>
</tr>
<tr>
<td>Not Engage</td>
<td>1.07</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.37)</td>
<td>(1.37)</td>
<td>(1.37)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Audience of Republicans &amp; President who is Democrat</th>
<th>Difference in Means</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Stay Out</td>
<td>0.61</td>
<td>-1.75</td>
<td>-1.36*</td>
</tr>
<tr>
<td>(0.50)</td>
<td>(0.50)</td>
<td>(-1.68)</td>
<td>(-1.85)</td>
</tr>
<tr>
<td>Engage</td>
<td>-2.36</td>
<td>-1.97</td>
<td></td>
</tr>
<tr>
<td>(0.66)</td>
<td>(-1.66)</td>
<td>(-1.63)</td>
<td>(-1.63)</td>
</tr>
<tr>
<td>Not Engage</td>
<td>0.39</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.39)</td>
<td>(0.39)</td>
<td>(0.39)</td>
</tr>
</tbody>
</table>

**Note:** *p<0.1; **p<0.05; ***p<0.01

Table B.3 shows the difference in average approval scores (Column Score – Row Score) and t-statistics among Democrats and Republicans based on strategies and the party of the president while holding territorial and casualty outcomes constant at no casualties and 20% territory gained by the attacking country. The approval score is measured on a scale from -3=Strongly Disapprove to 3=Strongly Approve.
Table B.4: Experiment 2: Differences in Approval Score for Strategies and Party

<table>
<thead>
<tr>
<th>Audience of Democrats &amp; President who is Democrat</th>
<th>Difference in Means</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stay Out</td>
<td>Engage</td>
</tr>
<tr>
<td>Audiences of Democrats &amp; President who is Democrat</td>
<td>-0.08 (-0.24)</td>
</tr>
<tr>
<td>Engage</td>
<td>-1.35*** (-3.97)</td>
</tr>
<tr>
<td>Not Engage</td>
<td>0.82*** (2.78)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Audience of Democrats &amp; President who is Republican</th>
<th>Difference in Means</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stay Out</td>
<td>0.03 (0.10)</td>
</tr>
<tr>
<td>Engage</td>
<td>-1.70*** (-5.00)</td>
</tr>
<tr>
<td>Not Engage</td>
<td>0.76*** (2.61)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Audience of Republicans &amp; President who is Republican</th>
<th>Difference in Means</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stay Out</td>
<td>0.97 (1.55)</td>
</tr>
<tr>
<td>Engage</td>
<td>-2.78*** (-4.61)</td>
</tr>
<tr>
<td>Not Engage</td>
<td>1.50*** (3.16)</td>
</tr>
</tbody>
</table>

Note: *p<0.1; **p<0.05; ***p<0.01

Table B.4 shows the difference in average approval scores (Column Score – Row Score) and t-statistics among Democrats and Republicans based on strategies and the party of the president while holding territorial and casualty outcomes constant at no casualties and 20% territory gained by the attacking country. The approval score is measured on a scale from -3=Strongly Disapprove to 3=Strongly Approve.
### B.6 Effects Among Independents

Table B.5: Experiment 1: Differences in Approval Score for Strategies and Party

<table>
<thead>
<tr>
<th>Audience of Independents &amp; President who is Democrat</th>
<th>Difference in Means</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Stay Out</td>
<td>Engage</td>
<td>Not Engage</td>
<td>Compromise</td>
</tr>
<tr>
<td>Stay Out</td>
<td>-1.12*</td>
<td>-1.63***</td>
<td>-0.85**</td>
</tr>
<tr>
<td></td>
<td>(-1.96)</td>
<td>(-3.28)</td>
<td>(-2.05)</td>
</tr>
<tr>
<td>Engage</td>
<td>-0.51</td>
<td>0.28</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-0.85)</td>
<td>(0.52)</td>
<td></td>
</tr>
<tr>
<td>Not Engage</td>
<td>0.79*</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.75)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Audience of Independents &amp; President who is Republican</th>
<th>Difference in Means</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Stay Out</td>
<td>-1.11*</td>
<td>-2.75***</td>
<td>-1.49***</td>
</tr>
<tr>
<td></td>
<td>(-1.95)</td>
<td>(-6.53)</td>
<td>(-3.71)</td>
</tr>
<tr>
<td>Engage</td>
<td>-1.64***</td>
<td>-0.38</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-3.00)</td>
<td>(-0.72)</td>
<td></td>
</tr>
<tr>
<td>Not Engage</td>
<td>1.26**</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(3.38)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note:* $^*p<0.1; ^{**}p<0.05; ^{***}p<0.01$

Table B.5 shows the difference in average approval scores (*Column Score* − *Row Score*) and t-statistics among Independents based on strategies and the party of the president while holding territorial and casualty outcomes constant at no casualties and 20% territory gained by the attacking country. The approval score is measured on a scale from -3=Strongly Disapprove to 3=Strongly Approve.
B.7 Mediation Analysis - Reputation and Support for Compromise

Figure B.2: Average Causal Mediation Analysis: The Effect of Reputation on Support for Compromise

Figure B.2 shows the average causal mediation effect (ACME) of reputation on audience approval, the average direct effect (ADE) of the Compromise treatment (relative to Not Engage), the total effect of both the mediation and treatment effects, and 95 percent confidence intervals, calculated using the mediation package in R (Tingley et al., 2014). Outcomes are held constant with no casualties and 20% territory gained by the attacking country. The approval score is measured on a scale from -3=Strongly Disapprove to 3=Strongly Approve, and the Reputation score is measured from 1 to 5.
B.8 Experiment 2 - Treatments

Experiment 2 followed a similar format at experiment 1, but includes additional treatments to test how sensitive audiences are to crisis outcomes. Each of the four treatment conditions varies:

1) The strategy of the president
2) The partisanship of the president
3) The outcomes of the crisis

Within the Compromise condition, there is an additional treatment that varies whether the U.S. president or the leader of the foreign country proposed the compromise. The crisis outcome treatments, which vary both the territory gained and casualties, are designed with two goals in mind. The first is to have common support across all conditions, meaning that there is a common policy outcome across all treatments. This allows the effect of the other treatments to be examined while holding outcomes constant. This study follows Kertzer and Brutger (2014) and uses 20 percent of the territory gained by the attacking country and no casualties as the common outcome. The second goal of the treatment is to have the variation in outcomes follow realistic conditions, meaning that shifts away from the common policy outcomes reflects that the U.S. receives better outcomes when they use force than when they don’t. Additionally, the casualty treatments are limited to the Engage condition, since it is not realistic to have military casualties when the U.S. does not engage. For a further discussion on casualties, please see appendix, section A-3.

- Stay Out: “The U.S. president, who was a [Democrat or Republican], said that the United States would stay out of the conflict. The attacking country continued to invade and took control of [20 or 50] percent of the contested territory.”
Consistent with Experiment 1, in each of the remaining treatment conditions the U.S. escalated with a threat and then the respondents were randomly assigned to one of the following conditions.

Threat: “The U.S. president, who was a [Democrat or Republican], said that if the attacking country continued to invade, the United States military would immediately engage and attempt to push out the attacking country. The president sent troops to the region.”

- Compromise: “The attacking country continued to invade, but the president did not immediately engage. The [U.S. president or leader of the attacking country] proposed a settlement, which was agreed to by all parties, where the attacking country stopped its invasion and took control of [20 or 50] percent of the contested territory.”

- Engage: “The attacking country continued to invade and the U.S. president ordered the military to engage. [The U.S. did not lose any troops or The U.S. lost 1,000 troops] in the conflict and the conflict ended with the attacking country [not taking control of any of the contested territory or taking control of 20 percent of the contested territory].”

- Not Engage: “The attacking country continued to invade. The U.S. president ordered the military not to engage. The attacking country continued to invade and took control of [20 or 50] percent of the contested territory.”
Appendix C

Decomposing Audience Costs

Appendix

C.1 Dispositional Measures

Militant assertiveness

- The best way to ensure world peace is through American military strength. [Strongly agree, agree, neither agree nor disagree, disagree, strongly disagree]

- The use of military force only makes problems worse. [Strongly agree, agree, neither agree nor disagree, disagree, strongly disagree]

- Going to war is unfortunate, but sometimes the only solution to international problems. [Strongly agree, agree, neither agree nor disagree, disagree, strongly disagree]

International trust

- Generally speaking, would you say that the United States can trust other nations, or that the United States can’t be too careful in dealing with other nations? [The United States can trust other nations, The United States can’t be too careful]
National chauvinism

- How superior is the United States compared to other nations? [Vastly superior, very superior, not so superior, not at all superior]

- How many things about America make you ashamed? [Very many, many, not many, none]

Political ideology

- Thinking about politics these days, how would you describe your political viewpoint? [Very liberal, liberal, slightly liberal, moderate, slightly conservative, conservative, very conservative]

C.2 Sampling Methodology and Sample Characteristics

The online survey experiment was fielded in the spring of 2014 on a national American sample of 942 registered voters recruited by Survey Sampling International (SSI). SSI panels employ an opt-in recruitment method, after which panel participants are randomly selected for survey invitations, using population targets rather than quotas to produce a nationally diverse sample of registered voters (see Table C.1 for sample characteristics). The experiment was embedded in a larger, unrelated survey, and participants were unaware of the content of the survey when they chose to participate. Because of the recruitment technique, the sample is nationally diverse, although not a national probability sample; for other examples of recent political science articles using SSI samples, see Barker, Hurwitz, and Nelson (2008); Healy, Malhotra, and Mo (2010); Popp and Rudolph (2011); Kam (2012); Malhotra and Margalit (2010); Malhotra, Margalit, and Mo (2013); Berinsky, Margolis, and Sances (2014).
Table C.1: Survey Sample Characteristics

<table>
<thead>
<tr>
<th>Category</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gender</strong></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>0.536</td>
</tr>
<tr>
<td>Female</td>
<td>0.464</td>
</tr>
<tr>
<td><strong>Age</strong></td>
<td></td>
</tr>
<tr>
<td>18-29</td>
<td>0.118</td>
</tr>
<tr>
<td>30-44</td>
<td>0.256</td>
</tr>
<tr>
<td>45-64</td>
<td>0.425</td>
</tr>
<tr>
<td>65+</td>
<td>0.201</td>
</tr>
<tr>
<td><strong>Education</strong></td>
<td></td>
</tr>
<tr>
<td>High school or below</td>
<td>0.249</td>
</tr>
<tr>
<td>Some college</td>
<td>0.412</td>
</tr>
<tr>
<td>College/university</td>
<td>0.215</td>
</tr>
<tr>
<td>Graduate/professional school</td>
<td>0.123</td>
</tr>
<tr>
<td><strong>Race/Ethnicity</strong></td>
<td></td>
</tr>
<tr>
<td>Hispanic/Latino</td>
<td>0.102</td>
</tr>
<tr>
<td>Non-Hispanic White</td>
<td>0.719</td>
</tr>
<tr>
<td>African-American</td>
<td>0.115</td>
</tr>
</tbody>
</table>

C.3 Alternative Experimental Designs

There are at least three alternative ways we could attempt to disentangle inconsistency costs from belligerence costs. First, we could employ open-ended questions asking respondents after the fact why they chose to punish the leader (e.g. [Levendusky and Horowitz, 2012; Davies and Johns, 2013]). The difficulty with this strategy is what psychologists call “introspection bias.” Decades of research on motivated reasoning tells us that people fare quite poorly at explaining why they formed their judgments, because so many of our opinions are generated outside the realm of conscious awareness, such that conscious introspection tends to produce ex-post rationalization rather than actually shedding light on the processes through which our opinions are formed ([Nisbett and Wilson, 1977; Lodge and Taber, 2013]). Indeed, self-perception theory emphasizes the extent to which individuals tend to infer their own attitudes and mental states by observing their behavior like a third-party would, rather than by accessing them directly ([Bem, 1972]. Open-ended responses are often valuable, and
recent methodological developments are making them increasingly powerful (Roberts et al., 2014), but they are less useful in disentangling the logic of audience costs than we might wish if, as Lodge, McGraw, and Stroh (1989, 401) note, “people can often tell you how much they like or dislike a book, movie, candidate or policy but not be able to recount the specific whys and wherefores for their overall evaluation.”

Second, we could employ an experimental design where we supplement the traditional audience cost experiment with a questionnaire that evaluates participants’ ex-ante attitudes towards the intervention, determining whether preferences for consistency are moderated by preferences for the policy itself. The trouble here, however, is that as Figure 2 in the main text shows, the audience cost treatment effect already includes both considerations about consistency (−a) and considerations about the use of force (−f) within the double-barreled audience cost treatment; testing whether this −a − f compound treatment effect is moderated by attitudes towards the use of force does not give us any explanatory purchase towards identifying how much of this extant treatment effect comes from each consideration.

Third, we could simply take the existing audience cost experimental design — which, as we argue in the main text, is implicitly a 2 × 2 with two cells missing — and turn it into an explicit one, adding the two missing off-diagonal cells: an “Engage” condition in which the President threatens the use of force and follows through (e.g. George W. Bush going to war against Iraq in 2003), and an opposing cell in which the President announces the country

1 A similar, although slightly different approach is taken by Chaudoin (2014c). A creative investigation into the dynamics of trade policy preferences, Chaudoin presents a scenario in which the President is considering imposing import restrictions to protect an American company; participants are presented with an argument in favor and an argument against the restrictions, and for a third of the sample, the anti-restriction argument tells participants that protectionist trade barriers are in violation of previous free trade agreements the United States has signed, and thus opens the US up to being sued at the WTO. By estimating this “international agreement” treatment effect, and measuring participants’ prior trade policy preferences, he is able to estimate whether the effect of the international agreement treatment is moderated by policy preferences, effectively disentangling these two effects. Although well-served in the context of his paper, this approach is inapplicable for audience cost experiments derived from the canonical crisis bargaining model, in which the audience cost treatment effect is −a − f, since estimating whether policy preferences moderate the impact of −a − f doesn’t actually separate −a and −f. We discuss this point again when interpreting the results from Table C.2 in Appendix §6.
will stay out of the conflict, and ends up reversing course by intervening (e.g. Woodrow Wilson winning the 1916 Presidential election on the slogan “He kept us out of war”, only to declare war on Germany in April 1917). We could then analyze the experimental data like a traditional 2 × 2 factorial experiment, where one factor represents the threat of force, and the other represents inconsistency. The problem here is that the latter possibility — a leader reneging on a pledge not to intervene — is not one considered by audience cost theory or traditional crisis bargaining models, as a look at the game tree in Figure 2 in the main text makes clear. Although a theoretically interesting extension of audience cost theory — do publics punish leaders for backing down on commitments to peace, just as they do for backing down on commitments to fight? — its inclusion here raises some methodological concerns (described more technically in the section below), so we set it aside for future research.

C.4 Simple Effects Versus Main Effects

Recall the 2 × 2 presented in Figure 2 in the main text. The experimental design is of the form:

\[ y_{ijk} = \mu_{ij} + \epsilon_{ijk} \]  \hspace{1cm} (C.1)

where \( i \) refers to the level of factor \( \alpha \) \((i = 1, 2)\), \( j \) refers to the level of factor \( \beta \) \((j = 1, 2)\), \( k \) indexes the participants in each cell \((k = 1...n_{ij})\), and the errors are independent and normally distributed with a mean of zero and a variance of \( \sigma^2 \) \((y_{ijk} \sim N(\mu_{ij}, \sigma^2))\). As the

\footnote{As Clausen (1966) notes, Wilson was initially leery about the campaign motto when an early version of it made its way into Martin Glynn’s keynote address at the 1916 Democratic national convention, but by the fall was regularly lambasting the Republicans for their interventionist stance, as in his speech at Shadow Lawn on September 30, 1916, where he warned that “the certain prospect of the success of the Republican Party is that we shall be drawn in one form or other into the embroilments of the European war.”}
missing cell indicates, the design is an incomplete rather than fully-crossed factorial, in that
\( n_{ij} \neq 0 \) for three of the four potential factor-level combinations: \( \mu_{11} \) (the leader decides to stay out), \( \mu_{21} \) (the leader makes a threat and follows through), and \( \mu_{22} \) (the leader makes a threat and backs down).

In the analysis, we estimate the following quantities of interest:

1. The simple effect of threatening force: \( \mu_{21} - \mu_{11} \) (or, \( -f \) from the game tree)

2. The simple effect of inconsistency: \( \mu_{22} - \mu_{21} \) (or, \( -a \) from the game tree)

As noted above, these are simple effects, rather than main effects: \( \mu_{21} - \mu_{11} \) is the effect of threatening force conditional on being consistent, while \( \mu_{22} - \mu_{21} \) is the effect of being inconsistent conditional on threatening force. Because of the missing cell, we are unable to estimate the main effects of factors \( \alpha \) and \( \beta \). However, although it may be somewhat counterintuitive to experimentalist in political science accustomed to studying main effects in factorial experiments as a matter of default, the main effects of \( \alpha \) and \( \beta \) are not actually our quantities of interest in audience cost experiments. Let us define the main effect of threatening force (\( \alpha \)) as:

\[
\mu_{i*} = \frac{\mu_{21} - \mu_{11}}{2} + \frac{\mu_{22} - \mu_{12}}{2}
\]

(C.2)

and the main effect of inconsistency (\( \beta \)) as:

\[
\mu_{j*} = \frac{\mu_{22} - \mu_{21}}{2} + \frac{\mu_{12} - \mu_{11}}{2}
\]

(C.3)

\(^3\)For discussions both of the problems with estimating main effects in factorials with missing cells, and of best practices for estimating simple effects for this type of data, see Freund 1980, Yandell 1997, Oehlert 2000, 233-4; Collins, Dziak, and Li 2009, 207, and Milliken and Johnson 2009, chap 13-14.
Because these main effects include the values of a cell not represented by a node on the game tree (with expected value $\mu_{12}$), they average over a simple effect outside the scope of the model: the main effect of inconsistency, for example, is the average of the simple effect of reneging on a threat of force (which is a part of the audience cost model), and the simple effect of reneging on a commitment to nonintervention (which is not). In other words, although running a fully-crossed factorial would allow us to estimate these main effects (as well as their interaction effect), they are not in themselves our quantities of interest, and would serve as biased estimates of them, such that the incomplete factorial used here, rather than a deficiency in experimental design, directly corresponds to our theory being tested.

C.5 The Impact of Specifying Casualty Levels on the Cost of Threatening Force

Even if one were not interested in decomposing the audience cost treatment into its constituent parts, in order to properly interpret the treatment effect, one would want to ensure that the treatment manipulated the President’s strategy while holding outcomes constant across conditions. After all, if the treatment effect compares a situation in which the President stays out and the US enjoys a serene status quo, with a situation in which the President backs down on a threat to go in and the country faces negative consequences, one is no longer estimating the effect of the President’s strategy on public approval, but rather the combined effect of the President’s strategy and its consequences. This is especially a concern, though, when we include the “Engage” treatment, since the traditional crisis bargaining model includes a war payoff ($w_1$) that assumes that the expected payoff of going to war is different from that of the other terminal nodes of the game tree. We avoid this problem by ensuring that participants are presented with an identical outcome (the conflict ends with the hostile state gaining 20 percent of the contested territory) regardless of whether the President stays
out, engages, or does not engage, such that changes in public approval can be traced entirely to the leader’s strategy.

Nonetheless, this raises a subtle point about experimental design, since even when faced with identical statements across vignettes, participants bring their own mental models to the table that may affect their perception of the payoffs: by specifying the US goes to war, participants may assume higher costs of war even if the outcomes are actually held constant across conditions (for a related point, see Sher and McKenzie (2006) and Tomz and Weeks (2013) on “information leakage”, as well as Dafoe, Zhang, and Caughey (2014)). Thus, in the analysis presented in the main text, we explicitly specify in the “Engage” condition that the US does not experience any casualties: this does not make the experimental conditions lexically equivalent (we do not mention casualties in the “Stay Out” and “Not Engage” conditions, since the US cannot experience casualties in an intervention that never takes place), but renders them logically equivalent. To test the consequences of this maneuver, we also ran an alternative version of our study in which participants in the “Engage” condition were not provided any information about casualties, rendering all three conditions lexically equivalent. When casualty conditions in the “Engage” condition are not specified, the belligerence cost estimate increases in size by 0.37 points (bootstrapped p < 0.055).

We suggest five considerations on this point. First, explicitly specifying zero casualties in the “Engage” condition thus constitutes a more conservative test for the existence of belligerence costs: belligerence costs account for 33% of the total audience cost when casualties are explicitly specified to be zero in the “Engage” condition, but 60% of the total audience cost when the conditions are lexically equivalent. Second, these casualty levels are artificially low; in the real world where we cannot separate policies from their consequences, few military interventions are as successful in terms of a loss of human lives. This suggests that we should think about the results presented in the main analysis as underestimating the magnitude of belligerence costs. Third, the fact that belligerence costs are higher when no
casualties are specified than when the US does not suffer any casualties means that we can reject an alternative interpretation of the main results, a “weak effort” argument in which the lack of casualties is taken as a sign that the President failed to exert sufficient effort in the intervention. Fourth, we can estimate a series of difference-in-differences modeling the change in the magnitude of the belligerence cost between different segments of the public to observe who is the most sensitive to not explicitly mentioning casualties; as one would expect, doves increase their belligerence costs more than hawks, cosmopolitans more than nationalists, and so on.

Finally, it is worth noting that we explicitly controlled for the human costs of war (Gelpi, Peaver, and Reifler 2009), but did not discuss other forms of costs such as economic costs (Geys 2010). Prominent research in international security emphasizes the extent to which casualties loom the largest as the “dominant indicator” people use when they think about the costs of war (Mueller 1971; Gartner 1997); indeed, Boettcher and Cobb (2006) find the American public to be extremely uncomfortable providing acceptable casualty levels for military scenarios, reflecting the extent to which human life is understood as sacred in a way that money is not (Tetlock 2003). Given the “information leakage” we document above, it is not clear a priori that explicitly controlling for economic costs would decrease the belligerence cost, since participants being told that a mission was unusually successful and painless in terms of human lives are likely to assume it was less costly in other respects as well. We therefore suggest that by controlling for casualties and conflict outcomes we are accounting for the largest consideration in people’s minds when they think about military

---

4 An IR literature on the political economy of war suggests the opposite hypothesis, that democratic publics understand capital and labor as substitutes (Gartzke 2001; Caverley 2009/10). Even if publics prefer capital-intensive wars based on firepower to labor-intensive wars based on manpower (one reason why the great powers have historically fared poorly at fighting insurgencies (Lyall 2010)), we expect that the public will tend to assume that the human and financial costs of war are correlated, a consequence of quagmires like Iraq that have proven to be as devastating in financial terms (Bilmes and Stiglitz 2012) as human ones.
engagement, and leave an investigation of how other factors shape belligerence cost estimates for future work.

C.6 Regression Analysis

In lieu of complex statistical models, the analysis in the main text calculates the quantities of interest using simple differences in means, exploring treatment-level heterogeneity simply by estimating a series of subgroup analyses, where we split the sample in two (separating the bottom quartile of respondents on a particular scale from the top quartile) and decompose the audience cost in each subgroup, estimating the fraction of the audience cost effect that comes from the costs of being consistent versus the costs of threatening force. Although this approach has the virtue of simplicity and transparency — we simply split the sample along different “axes” and test how the logics of audience costs change — it does not control for multiple dispositional characteristics simultaneously, as would be done in a conventional regression model. Since a number of these dispositional traits are correlated with one another — 50% of respondents in the bottom quartile of national attachment also score in the bottom quartile of national chauvinism, for example — it is worth controlling for multiple dispositional traits simultaneously to ensure that the effects posited to stem from one of these characteristics are actually not stemming from another.

These factors include, for example, opportunity costs, the intrinsic value of fighting independent of the outcome, the purpose of the intervention, prior interactions with the hostile state, and so on.
<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
</tr>
<tr>
<td>Not Engage</td>
<td>1.754***</td>
<td>-1.247***</td>
<td>-1.098***</td>
<td>-1.001***</td>
</tr>
<tr>
<td></td>
<td>(0.384)</td>
<td>(0.214)</td>
<td>(0.369)</td>
<td>(0.419)</td>
</tr>
<tr>
<td>Engage</td>
<td>1.615***</td>
<td>-0.118</td>
<td>-0.487</td>
<td>-0.792</td>
</tr>
<tr>
<td></td>
<td>(0.431)</td>
<td>(0.244)</td>
<td>(0.442)</td>
<td>(0.497)</td>
</tr>
<tr>
<td>Militant Assertiveness</td>
<td>0.726*</td>
<td>-0.576</td>
<td>-0.655</td>
<td>-0.671</td>
</tr>
<tr>
<td></td>
<td>(0.387)</td>
<td>(0.467)</td>
<td>(0.553)</td>
<td>(0.670)</td>
</tr>
<tr>
<td>International Trust</td>
<td>0.820***</td>
<td>1.139***</td>
<td>0.357</td>
<td>1.048***</td>
</tr>
<tr>
<td></td>
<td>(0.292)</td>
<td>(0.318)</td>
<td>(0.263)</td>
<td>(0.325)</td>
</tr>
<tr>
<td>National Chauvinism</td>
<td>0.632</td>
<td>0.711</td>
<td>0.485</td>
<td>0.262</td>
</tr>
<tr>
<td></td>
<td>(0.581)</td>
<td>(0.474)</td>
<td>(0.354)</td>
<td>(0.683)</td>
</tr>
<tr>
<td>Ideology</td>
<td>-0.076</td>
<td>-0.054</td>
<td>-0.067</td>
<td>0.164</td>
</tr>
<tr>
<td></td>
<td>(0.068)</td>
<td>(0.055)</td>
<td>(0.066)</td>
<td>(0.461)</td>
</tr>
<tr>
<td>Not Engage × Mil Assert</td>
<td>0.549</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Engage × Mil Assert</td>
<td>1.858***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.497)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not Engage × Intl Trust</td>
<td></td>
<td>0.265</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Engage × Intl Trust</td>
<td></td>
<td>-1.394***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.542)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not Engage × Nat Chauv</td>
<td></td>
<td></td>
<td>0.532</td>
<td></td>
</tr>
<tr>
<td>Engage × Nat Chauv</td>
<td></td>
<td></td>
<td>0.005</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.481)</td>
<td></td>
</tr>
<tr>
<td>Not Engage × Ideology</td>
<td></td>
<td></td>
<td>0.558</td>
<td></td>
</tr>
<tr>
<td>Engage × Ideology</td>
<td></td>
<td></td>
<td>0.037</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.699)</td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>-0.677***</td>
<td>-0.467**</td>
<td>-0.704***</td>
<td>-0.530</td>
</tr>
<tr>
<td></td>
<td>(0.287)</td>
<td>(0.212)</td>
<td>(0.261)</td>
<td>(0.344)</td>
</tr>
<tr>
<td>Male</td>
<td>1.001</td>
<td>-0.084</td>
<td>-0.219</td>
<td>-0.141</td>
</tr>
<tr>
<td></td>
<td>(0.223)</td>
<td>(0.170)</td>
<td>(0.212)</td>
<td>(0.268)</td>
</tr>
<tr>
<td>Income</td>
<td>-0.121*</td>
<td>-0.084</td>
<td>-0.089</td>
<td>-0.119</td>
</tr>
<tr>
<td></td>
<td>(0.068)</td>
<td>(0.052)</td>
<td>(0.065)</td>
<td>(0.079)</td>
</tr>
<tr>
<td>Education</td>
<td>-0.075</td>
<td>-0.038</td>
<td>-0.056</td>
<td>-0.020</td>
</tr>
<tr>
<td></td>
<td>(0.051)</td>
<td>(0.039)</td>
<td>(0.048)</td>
<td>(0.061)</td>
</tr>
<tr>
<td>Constant</td>
<td>1.984***</td>
<td>1.284***</td>
<td>2.062***</td>
<td>1.364***</td>
</tr>
<tr>
<td></td>
<td>(0.590)</td>
<td>(0.451)</td>
<td>(0.554)</td>
<td>(0.673)</td>
</tr>
<tr>
<td>N</td>
<td>368</td>
<td>588</td>
<td>390</td>
<td>271</td>
</tr>
<tr>
<td>Adjusted R²</td>
<td>0.430</td>
<td>1.241</td>
<td>0.108</td>
<td>0.118</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Inconsistency Fraction</th>
<th>Low Mil Assert: 0.044***</th>
<th>Low Int Trust: 0.921***</th>
<th>Low Nat Chauv: 0.569</th>
<th>Liberals: -0.207</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low Mil Assert: 1.27***</td>
<td>High Int Trust: 0.075***</td>
<td>High Nat Chauv: 0.711</td>
<td>Conservatives: 0.737</td>
<td></td>
</tr>
</tbody>
</table>

*p < .1; **p < .05; ***p < .01; Note that the quantities of interest here are not the regression coefficients but the inconsistency fractions calculated in the bottom two rows of the table, and their associated p-values derived from joint hypothesis tests. For ease of interpretation, we recommend comparing results visually using Figure C.1.
We thus also adopt a regression framework, estimating a series of OLS models of the form:

$$y = \beta_0 + \beta_1 T_{22} + \beta_2 T_{21} + \beta_3 Z + \beta_4 T_{22}Z + \beta_5 T_{21}Z + \beta_6 x + \ldots + \beta_j x + \epsilon$$  \hspace{1cm} (C.4)

in which $T_{22}$ corresponds to the treatment with expected value $\mu_{22}$, $T_{21}$ represents the treatment with expected value $\mu_{21}$, $z$ is a dispositional variable of interest, $\beta_4 T_{22}Z$ and $\beta_5 T_{21}Z$ are interaction terms between the dispositional variable and the two treatments, and $\beta_6 x \ldots \beta_j x$ represent a series of dispositional control variables.\footnote{We are grateful to Yusaku Horiuchi for helpful discussions on this section. The regression table in Table C.2 is partially generated using the Stargazer package in R \cite{Hlavac2013}.} The remaining treatment, $T_{11}$ (with expected value $\mu_{11}$), is the reference category, and thus not included in the model. To facilitate a direct comparison between the regression analyses and the subgroup analyses employed in the main text, $Z$ is dichotomized, such that $Z = 0$ for the bottom quartile of respondents along a particular trait, and $Z = 1$ for the top quartile, and the middle two quartiles are...
Our quantities of interest can then be expressed as:

\[
E[Y_{22} - Y_{11}|Z = z] = \beta_1 + \beta_4 Z \quad \text{(C.5)}
\]

\[
E[Y_{21} - Y_{11}|Z = z] = \beta_2 + \beta_5 Z \quad \text{(C.6)}
\]

\[
E[Y_{22} - Y_{21}|Z = z] = (\beta_1 - \beta_2) + (\beta_4 - \beta_5) Z \quad \text{(C.7)}
\]

\[
\frac{E[Y_{22} - Y_{21}|Z = z]}{E[Y_{22} - Y_{11}|Z = z]} = \frac{\beta_1 - \beta_2}{\beta_1 + \beta_4} + \frac{\beta_4 - \beta_5}{\beta_1 + \beta_4} Z \quad \text{(C.8)}
\]

\[
\frac{E[Y_{21} - Y_{11}|Z = z]}{E[Y_{22} - Y_{11}|Z = z]} = \frac{\beta_2 + \beta_5 Z}{\beta_1 + \beta_4} Z \quad \text{(C.9)}
\]

We therefore derive our parameter estimates from 2000 bootstrapped versions of each of our regression models to derive probability distributions for each of our quantities of interest. Table C.2 displays coefficients and standard errors from a series of regression models. The number of interaction terms means that these regression models are slightly less transparent than the simple mean-split results reported in the main paper, but careful inspection reveals that they replicate the findings from Table 1 of the main text. For example, the first interaction term in model 1, Not Engage × Mil Assert, shows that hawks and doves do not demand audience costs of significantly different sizes. However, although the magnitude of the total audience cost remains the same, the logic behind the audience cost differs: the significant Engage × Mil Assert interaction shows that hawks and doves demand belligerence costs of significantly different sizes – a finding that would be masked if we simply looked only at the total audience cost. Thus, these regression results emphasize the importance of disaggregating the audience, given that a traditional analysis of audience cost treatment effects would erroneously lead a reader to believe that audience characteristics are insignificant when determining audience reactions.

---

7Thus, to avoid further reductions in sample size, all of the dispositional controls in \(\beta_6 x \ldots \beta_j x\) are kept as continuous. Substantively similar results are obtained when \(Z\) is treated as continuous rather than dichotomized.

8In this sense, the results of this table reinforce the claim made in Appendix §3 about how measuring participants’ policy preferences and seeing whether they moderate the magnitude of the audience cost treatment effect is inadequate at disentangling the two logics of audience costs.
Most importantly, the bottom section of each table depicts our main quantity of interest
the fraction of the audience cost treatment that comes from inconsistency \((E[Y_{22} - Y_{21} | Z=z] / E[Y_{22} - Y_{11} | Z=z])\)
(which is a complement of our other quantity of interest, the fraction of the audience cost
treatment that comes from threatening force \((E[Y_{21} - Y_{11} | Z=z] / E[Y_{22} - Y_{11} | Z=z])\)); we employ bootstrapped joint
hypothesis tests to calculate whether the inconsistency cost fraction significantly differs
across each of the two subgroups. The key conclusion to draw from Table \text{[C.2]} comes from
comparing the inconsistency fractions calculated in the bottom two rows with their counter-
parts in Table 1 of the main text: the estimates produced by the simple subgroup analyses
in the main text do not substantively differ from those that employ a more complicated
regression approach, further confirmed by directly comparing the probability distributions
for each approach in Figure \text{[C.1]}

C.7 Alternative Measure of Approval

While our main analysis focuses on respondents’ change in approval measured along a seven-
point scale, other studies have examined audience costs using percentages instead, examining
the percent of the audience who disapproves of the president (\cite{Tomz2007}). To enhance com-
parability to previous studies and to highlight the magnitude of the changes in the percent
of the audience punishing the President’s handling of the situation, Table \text{[C.3]} presents our
results based on changes in the percent of respondents who strongly or somewhat disapprove
of the President’s handling of the situation. When compared to the results in Table 1 of the
paper, it is clear that belligerence and inconsistency costs follow the same pattern regardless
of which measurement is used. Examining changes in the percent of the audience who
punishes the leader also highlights the magnitude of the changes in approval that are driven
by both belligerence and inconsistency costs. For example, among doves, the belligerence
cost leads to a 31 percentage point increase in the number of audience members who disap-
prove, whereas inconsistency leads to a 10 percentage point increase in those who disapprove

148
among doves. The magnitude of the results reported in Table C.3 demonstrate that both belligerence and inconsistency can have great costs to the president, depending on who the President’s relevant audience is.

Table C.3: Decomposing Audience Costs (In Percentage Point Change in Disapproval)

<table>
<thead>
<tr>
<th>Audience</th>
<th>Cost (%)</th>
<th>Belligerence</th>
<th>Inconsistency</th>
<th>Belligerence Cost Fraction</th>
<th>Inconsistency Cost Fraction</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
<td>(5)</td>
</tr>
<tr>
<td></td>
<td>$\mu_{22} - \mu_{11}$</td>
<td>$\mu_{21} - \mu_{11}$</td>
<td>$\mu_{22} - \mu_{21}$</td>
<td>$\frac{\mu_{21} - \mu_{11}}{\mu_{22} - \mu_{11}}$</td>
<td>$\frac{\mu_{22} - \mu_{21}}{\mu_{22} - \mu_{11}}$</td>
</tr>
<tr>
<td>Full Sample</td>
<td>+28.31</td>
<td>+7.67</td>
<td>+20.65</td>
<td>0.27</td>
<td>0.73</td>
</tr>
<tr>
<td>Low Militant</td>
<td>+41.15</td>
<td>+31.47</td>
<td>+9.68</td>
<td>0.76</td>
<td>0.24</td>
</tr>
<tr>
<td>High Militant</td>
<td>+25.21</td>
<td>-4.36</td>
<td>+29.57</td>
<td>-0.17</td>
<td>1.17</td>
</tr>
<tr>
<td>Low Int Trust</td>
<td>+27.76</td>
<td>+2.24</td>
<td>+25.52</td>
<td>0.08</td>
<td>0.92</td>
</tr>
<tr>
<td>High Int Trust</td>
<td>+28.86</td>
<td>+26.57</td>
<td>+2.3</td>
<td>0.92</td>
<td>0.08</td>
</tr>
<tr>
<td>Low Nat Chauv</td>
<td>+26.72</td>
<td>+12.78</td>
<td>+13.94</td>
<td>0.48</td>
<td>0.52</td>
</tr>
<tr>
<td>High Nat Chauv</td>
<td>+33.24</td>
<td>+8.95</td>
<td>+24.29</td>
<td>0.27</td>
<td>0.73</td>
</tr>
<tr>
<td>Liberals</td>
<td>+28.22</td>
<td>+19.99</td>
<td>+8.24</td>
<td>0.71</td>
<td>0.29</td>
</tr>
<tr>
<td>Conservatives</td>
<td>+31.69</td>
<td>+1.82</td>
<td>+29.86</td>
<td>0.06</td>
<td>0.94</td>
</tr>
</tbody>
</table>

Note: Columns 1-3 depict treatment effects calculated using the percentage point change in respondents who strongly or somewhat disapprove of the president’s handling of the situation. For example, 28.31 percentage points more of the audience in the full sample disapproves of the President’s handling of the crisis when she backs down on a threat than when she stays out in the first place. Columns 4-5 calculate the fraction of the total audience cost stemming from threatening force (4), versus from inconsistency (5). Note that these two fractions add up to 1.
Bibliography


Political Accountability and Concessions in Crisis Diplomacy.” *Security Studies* 20 (2): 
141-170.


University Press.


Cavari, Amnon. 2013. “Religious Beliefs, Elite Polarization, and Public Opinion on Foreign 


Chaudoin, Stephen. 2014b. “Promises or Policies? An Experimental Analysis of International 

Chaudoin, Stephen. 2014c. “Promises or Policies? An Experimental Analysis of International 


Trade Attorney. 2009. Author Interview. Washington, DC.

Trade Official. 2014. Author Correspondence. Washington, DC.


Trubowitz, Peter, and Nicole Mellow. 2005. ““Going Bipartisan”: Politics by Other Means.” *Political Science Quarterly* 120 (Fall): 433-453.


USTR General Counsel. 2009. Author Interview. Washington, DC.

USTR Official. 2009. Author Interview. Washington, DC.


