Emotion and Coercive Credibility in International Crises

by

Hohyun Yoon

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Political Science)

at the

UNIVERSITY OF WISCONSIN-MADISON

2024

Date of Final Oral Examination: 07/08/2024

The dissertation is approved by the following members of the Final Oral Committee:

Andrew H. Kydd, Professor, Political Science (chair)
Jon C.W. Pevehouse, Professor, Political Science/La Follette School of Public Affairs
Jessica L.P. Weeks, Professor, Political Science
Jonathan Renshon, Professor, Political Science
Adeline Y. Lo, Assistant Professor, Political Science

Emotion and Coercive Credibility in International Crises

Hohyun Yoon

Abstract

Why are some threats more effective than others at influencing adversaries' beliefs and the outcomes of international crises? This dissertation aims to answer this question by focusing on the threats' credibility. More specifically, I propose elite expressions of anger as a key determinant of the credibility of coercive threats. Most existing studies that seek to explain the credibility of threats highlight factors that shape the costs of making or backing down from threats, but largely overlook the content of the statements leaders make. In this dissertation, I argue that leader anger is an essential condition for threat credibility and describe two specific mechanisms. Leaders conveying anger can project an image of resolve as they are seen as less sensitive to the cost of fighting, and anger expressions increase the reputational consequences of threats.

Across three empirical chapters, I show that threats accompanied by higher levels of anger expressions are more credible. In doing so, I offer a novel dataset on public statements of threats and anger expressions and threats, built by hand-labeling world leaders' public speeches made in international crises that took place between 1946-1996. In addition, through randomized experiments, I show that an adversary's anger expressions increase observers' beliefs about the adversary's willingness to use force, and that this effect is mediated by the observers' beliefs about the adversary's insensitivity to the cost of fighting. I also find evidence that a leader's anger expressions increase the domestic public's concerns about the country's reputation. My dissertation provides a new explanation for threat credibility and a methodologically innovative empirical study of emotions in international relations.

Dedication

I dedicate my dissertation to my wife, Hyeyoung Koh, and my parents, Ki-eun Yoon and Hyo-sin Lee.

Acknowledgements

First and foremost, I would like to thank my fantastic advisors and members of my dissertation committee, Andy Kydd, Jon Pevehouse, Jessica Weeks, Jonathan Renshon, and Adeline Lo. Andy was open to different theoretical and methodological perspectives. I changed my dissertation topic twice (!)—from nuclear weapons to alliances, and then to emotions in international crises—and Andy was very patient with me throughout the process and provided crucial insights that eventually led me to write this dissertation. Jon and Jessica have been excellent mentors to me, in terms of both career and research. They were very generous with their time and always willing to help me not just on my dissertation work but also on various other things I needed as a grad student to get through the graduate school. Jonathan was the reason why I was able to do experimental research. He was helpful at all stages of the experimental work that I conducted for my dissertation. If it weren't for him, the dissertation would only have one empirical chapter. Adeline was always available to discuss my work, and I was extremely fortunate to work with her through years of research assistantship. I cannot express how thankful I am for her time and support. Again, all of my mentors were closely involved in my academic journey, and I am so grateful.

I am also thankful for the help I received from people outside the Political Department at the University of Wisconsin. I thank Burcu Savun for her careful comments on my job market paper, and Michaela Mattes, Santiago Olivella, Eric Min, Max Goplerud, Roseanne McManus, Danielle Jung, and Christopher Fariss for their advice on the job market.

I am also grateful to my friends and colleagues. In particular, I'd like to thank Andrew McWard and Yumi Park for being good friends and inspiring coauthors. I published my first journal article with Andrew, and he continues to impress me with new interesting research ideas. Yumi was a second-year student when I came to Madison. She never said no whenever I pestered her with questions or asked for help. I am very lucky to have great friends like them.

I am incredibly thankful for the continued support of my parents, Ki-eun Yoon and Hyo-sin Lee, and my sister, Young-sun Yoon. I know political science might not be the most exciting topic for family gatherings, but they've always shown genuine interest in my work. Even though I've been living abroad for over ten years, your constant encouragement has made me feel never alone. Thank you, I love you all.

Lastly, and most importantly, I thank Hyeyoung for her endless love and

support. Finishing the doctoral degree wouldn't have been possible without her. We met as graduate students in Madison, and now we're leaving together as doctors. I'm so excited for what the future holds for us in Pullman!

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Coercive diplomacy is the art of forcefully persuading an opponent to change its behavior. Scholars of international politics have long recognized coercive diplomacy as a peaceful alternative to war, in which a state seeks to coerce an opponent into conceding a disputed good by demonstrating its ability and willingness to use force rather than actually using force to take it from the opponent (George 1991; Schelling 1960). Of many strategies of coercion available, verbal threats to use force are thought to be a particularly efficient tool because, if successful, they allow the state to achieve its aims without bearing the cost of taking material actions (Fearon 1997).

However, threats are not always believed. Explaining the conditions under which threats are credible has thus been a central research topic in the field of International Relations (IR). We now understand a great deal about the role of costly signaling (Fearon 1997; Slantchev 2005), domestic audience costs and regime type (Fearon 1994; Schultz 1998; Tomz 2007*a*; Gelpi and Griesdorf 2001), international reputations and strategic considerations (Sartori 2002; Guisinger and Smith 2002; Trager 2010), and private diplomacy (Kurizaki 2007; Yarhi-Milo 2014; Holmes 2013).

Yet, despite generations of rich scholarship, none of the debates on these explanations have been fully resolved. A number of scholars question whether costly signals are perceived as intended (Jervis 1976; Quek 2016; Kertzer, Rathbun and Rathbun 2020), audience costs exist (Snyder and Borghard 2011; Trachtenberg 2012; Levendusky and Horowitz 2012; Croco, Hanmer and McDonald 2021), and regime type matters (Downes and Sechser 2012; Weeks 2008). There is limited empirical work on the relative effectiveness of public and private threats (Katagiri and Min 2019; Trager 2017), and the enormous literature on reputations has not yet agreed on whether and how reputations emerge and influence international crises (Press 2005; Mercer 1996; Hopf 1994; Harvey and Mitton 2016). In sum, we still do not fully understand the precise effects and mechanisms of what makes threats credible.

Moreover, most of the existing literature focuses on contextual factors in which the speech acts occur. Only a few studies examine threats themselves empirically (Katagiri and Min 2019), and fewer studies explicitly theorize about variation *within* statements (Trager 2017; McManus 2017).[1] This is problematic because adversaries likely draw inferences about each other's intentions not only from contextual factors but also from the content of the statements. By leaving out threats, theories about threat credibility will inevitably have limited explanatory power. For example, explanations based on power and regime type are hard-pressed to explain why between 1918-2001, compellent threats from three democratic great powers—the United States, the U.K., and France—succeeded only 36% of the time combined, whereas those from the Soviet Union succeeded at a 58% rate (Sechser 2011). Moreover, relatively unchanging power and regime type cannot explain why U.S. threats to intervene in Kuwait were ignored by Iraq in 1990 but believed

[1] Indeed, one major criticism against the audience cost literature is that most of the crises and disputes data used to provide evidence for threat credibility do not contain explicit threats (Downes and Sechser 2012).

in 1994, only to be ignored again in 2003.

To address these gaps, this dissertation develops and empirically tests a theory about variation within diplomatic statements issued by political leaders. I focus on a particular source of variation: the emotionality of leaders. Drawing in social and evolutionary psychology, I propose leader anger conveyed in threats as a crucial mechanism for credibility in coercive diplomacy. I show that leaders increase the credibility of their threats by revealing anger, that the effects of anger operate through the perceptions that the leader is insensitive to the cost of fighting, and that leaders' anger expressions may have longer-term consequences for the country's reputation for making honest threats. Studying anger expressions adds to a more comprehensive understanding of threat credibility by offering a leader-level explanation. It provides leverage to account for variation among states in similar strategic and political environments, as well as variation among leaders within the same state and over time for the same leader.

1.1 Existing Insights about Threat Credibility

In the traditional model of coercive bargaining, the coercer, dissatisfied with the status quo, sends a threat to the target with hopes of extracting a concession. The target then decides whether to concede or resist. If the target concedes, the bargaining ends. If it resists, the coercer either carries out the threat or backs down. In order for the target to concede, it must first and foremost believe that the coercer will punish it if it resists.

How can the target know that it will be punished? The vast majority of the scholarship on coercive diplomacy points to *credibility* as the crucial factor. The credibility of a threat is the extent to which an actor is perceived to have the

capability to carry out its threat and the willingness to fight.[2] Capability may be an obvious factor since it determines the threatening state's ability to follow through (McManus 2017; Press 2005). The picture for the willingness to fight, or resolve, is more complicated. Resolve depends on the actor's valuation of the issue at stake and the cost of fighting as well as individual dispositions (Mercer 1996; Kertzer 2016). The fundamental problem is that all of these factors—interests, the cost of fighting, dispositions, and even capability—are to a large extent private information, meaning that the threatening state always knows more about them than the target. The problem is compounded by the fact that adversaries in international crises have strategic incentives to exaggerate their resolve by making empty threats (Fearon 1995). Without such incentives, credibility would be irrelevant as states will truthfully communicate their resolve. However, because each side understands its opponent's incentives to misrepresent, any state resolved to issue a genuine threat must overcome the challenge of demonstrating its credibility.

This problem leads to the influential rationalist solution that threats are credible when they incur costs to the sender. Rationalist scholars conceive of these costs as strategic costs that states or leaders suffer when they make threats or fail to fulfill them, independently of how much they value the stake at hand. Perhaps the most prominent theory in this literature is audience cost theory (Fearon 1994; Tomz 2007*a*). Audience costs refer to the punishment leaders suffer when they renege on foreign policy commitments that they have publicly made. The punishment is usually conceptualized as political disapproval from domestic actors that can threaten

[2]Credibility should be distinguished from resolve itself, although they are often used interchangeably in the literature. Resolve refers to an actor's general willingness to fight, whereas credibility is assigned to a specific threat. For example, a state can be resolved without making any threats. A threat as a verbal declaration of intentions aims to influence a target's beliefs about the sender's resolve, but the threat has to be perceived as credible to exert any influence on resolve. Conversely, if a target already knows that the sender is resolved (e.g. states with strong reputations for resolve), then the threat will be credible almost by definition, and this will be an uninteresting case where the issue of credibility would be moot. In practice, however, the fact that resolve is private information means that no state can know for certain the other's inherent willingness to fight.

the leader's stay in power. The core argument is that publicly threatening to use force enables leaders to generate audience costs to backing down because domestic audiences, who are concerned about their country's international reputation and future credibility, punish leaders for called bluffs (Fearon 1997; Guisinger and Smith 2002; Smith 1998). The implication for international crises is that leaders, who generally wish to avoid audience costs, will only make public threats that they will follow through on. This domestic political dynamic gives a bargaining advantage to a leader who threatens to force in public rather than through private diplomacy.

Another implication of audience cost theory is that regime type matters. One key insight from this line of work is that a crucial factor determining threat credibility is the political vulnerability of the leader issuing threats. Advocates argue that democratic leaders are able to generate greater audience costs and thus make more credible threats than their autocratic counterparts because democratic institutions make it easier for the domestic public to punish leaders for making empty threats (Fearon 1995; Gelpi and Griesdorf 2001; Gelpi and Grieco 2015). Schultz (1998) argues that incumbent leaders can make their threats more credible when there is an independent opposition party than when the state is ruled by an authoritarian ruler or a single party. The opposition party can independently support the incumbent's threat in a crisis, and the decision to support the threat means the party is sufficiently certain that the threat is likely to succeed or the state is sufficiently resolved. In a similar vein, Potter and Baum (2014) suggest that democratic states with free, independent media are better able to issue credible threats than states where the media is constrained.

Threats are believed to create strategic costs at the international level as well. Scholars working on this topic propose ways in which diplomatic threats can be credible without being public. Trager (2010) argues that threats carry some

inherent strategic costs as the adversary could take actions that place the coercer at military disadvantage in respond to the threats. For example, the target might make war preparations, form hostile alliances, or launch a pre-emptive attack. (Sartori 2005) argues that leaders are discouraged from making threats they would not be willing to follow through because having their bluff called damages their countries' international reputation, jeopardizing future threat credibility.

1.2 How Do Targets Really Assess Credibility?

Yet, despite providing the clear logic of costly signaling, the literature focuses almost exclusively on the threatening states, assuming that the costly threats will be objectively interpreted as intended by the senders. However, scholars working in the tradition of political psychology have long recognized the problems of misinterpretation and misperception (Jervis 1976). Recent empirical studies suggest that observers do not always assess credibility from the theoretically prominent factors proposed by theories of costly signaling, such as regime type. Multiple studies across different methods fail to find evidence that threats from democracy are perceived as more credible than those from nondemocarcies or that targets use the threatening state's regime type to draw inferences about credibility (Trachtenberg 2012; Snyder and Borghard 2011; Yarhi-Milo, Kertzer and Renshon 2018; Sechser 2011; Downes and Sechser 2012). More generally, Quek (2016) shows in lab experiments that only senders, but not receivers, of costly threats behave in a way consistent with the predictions of costly signaling theories. Furthermore, several experimental findings suggest that assessing domestic audience costs is likely more difficult than scholars have portrayed it to be. This is because such assessments require understanding the audience's preferences and dispositions (Kertzer and Brutger 2016; Nomikos and Sambanis 2019; Chaudoin 2014). Military

power, another prominent variable, does not fare better. Examining a dataset on compellent military threats, Sechser (2011) suggests that for the period between 1918-2001 threats from militarily stronger states may not be more successful than threats from weaker countries. Echoing this observation, Art and Greenhill (2018) describe the United States' "underwhelming track record" of coercive success as puzzling.

This literature reveals a considerable gap in how actors perceive the credibility of a threat depending on which end of the signaling mechanism they stand on. The sender of a threat knows a lot more than the receiver about its own domestic audiences, political institutions, and many other contextual factors that shape the costs. If the credibility of the threat depends on these numerous contextual factors, of which the receiver is uncertain anyway, this will significantly weaken the effectiveness of threats even when a threatening leader and a target would agree on the principle of costly signaling. Of course, states cannot literally "calculate" credibility, and may instead use observable variables as indicators (Kertzer, Renshon and Yarhi-Milo 2021). Even then, the target will face a serious decision problem as there are simply too many countervailing indicators of intentions in a crisis situation. Inferring resolve from a threat thus likely remains what Kertzer, Renshon and Yarhi-Milo (2021) call "a computationally intractable, ill-defined decision problem, characterized by irreducible uncertainty."

This raises a question of what leaders can do to bridge the sender-target gap in coercive diplomacy. Without denying the significance of previous scholarly work, I argue that the literature has not paid sufficiently attention to the actual diplomatic communications. Given the influential role leaders play on shaping international events, a fuller account of the credibility of threats should consider what else leaders say in their statements of threats and how they say it. Some counterfactual considerations illustrate the point. Did the domestic and strategic

contexts make the failure of Obama's "red line" warning against Syria's use of chemical weapons unavoidable, or might an alternative rhetoric have changed the way his threat was perceived? Does Chinese's officials' explosive anger over the issue of Taiwan influence foreign observers' of China's resolve, or is the anger nothing more than noise? These questions, although difficult to answer, illustrate the importance of considering how rhetoric and style might matter in coercive diplomacy.

1.3 Why Emotional Expressions?

It is worth explaining why it is important to bring emotional expressions to the study of coercive diplomacy. Incorporating emotions has already generated valuable insights on pathways to conflict that may not be readily explained by rationalist theories of conflict (Kertzer and Tingley 2018; Hafner-Burton et al. 2017). We now understand much about how emotions affect individuals' foreign policy preferences (Herrmann 2017; Halperin, Canetti-Nisim and Hirsch-Hoefler 2009; Zeitzoff 2014; Powers and Renshon 2021), as well as information processing (Renshon, Lee and Tingley 2017). Several scholars also theorize about specific political processes by which emotions can lead to collective emotional experiences and influence state decisions (Hall and Ross 2015; Sasley 2011). However, despite claims that emotions "play an obvious and omnipresent role in world politics" (Bleiker and Hutchison 2008), our investigation of the topic tends to be limited to the domain of individual decision making. This analytic focus has left unanswered important questions about how we might integrate emotions into more conventional models of international politics characterized by strategic interactions. For example, if anger can cause an actor to be more resistant to coercion, what would a state do if it knows that its actions could make its adversary angry? What would

a state do if it knows that its opponent knows that anger would make it more resistant? How are signals from actors in anger perceived differently than actors in other emotional states? The literature on coercive diplomacy currently lacks an empirical foundation to answer these types of questions. While scholars of diplomacy have shed light on interpersonal emotions in the context of face-to-face diplomacy (Hall and Yarhi-Milo 2012), they have left other forms of interstate communication unexplored, and there are methodological limitations as they rely on qualitative case studies.

A solution I propose is to consider the socio-functional value of emotions. Individuals in world politics not only feel emotions but also communicate their emotions. A voluminous social psychological work shows that an actor's emotional state affects others' behavior toward that actor by conveying information about the actor's mental states and intentions (Keltner and Haidt 1999; Van Kleef 2009). This suggests that communicating emotions in crisis situations might influence adversaries in a way that is not accounted for by existing theories. Take anger as an example. Rationalist theories would assume that emotions are noise and predict that anger has no effect on an adversary's behavior. Behavioral theorists would argue that anger should foster conflict as an angry actor is likely to be hawkish. From the social perspective, however, anger from an actor could inform an adversary that the actor has become more inclined to using force, potentially increasing the probability that the adversary backs down. In sum, if emotions matter because they systematically influence individual attributes relevant to decisions over war and peace, emotion expressions should also matter because they could help outside observers make predictions about the actor's likely behavior (Van Kleef, De Dreu and Manstead 2010).

1.4 Methodological Approach

The focus on emotion expressions offers a considerable methodological advantage. First, the communication of emotions is much more observable than the experience of emotions. Since political scientists can rarely measure leaders' emotions in real time during a crisis, qualitative case studies and experiments have been the primary methods in the study of emotions in IR. While each of the two has its advantages and drawbacks, there have been concerns about the generalizability of findings to international outcomes from historical cases selected by the researcher or experiments that typically employ abstract vignettes and non-elite subject pools. Yet, we are much better able to observe and measure political elites' verbal and, perhaps to a lesser extent, nonverbal emotional communications. Methodological diversity is crucial for the study of emotions in world politics because emotions are elusive and cannot be measured directly. Focusing on expressions of emotions thus facilitates a multi-method research design that incorporates a large-N cross-national data, the approach this project adopts.

More specifically, qualitative case studies, the historically dominant method, have generated valuable insights about how emotions influenced leaders' foreign policy in a specific way by analyzing decision making processes and interpersonal interactions (e.g., Wong 2019; Hall and Yarhi-Milo 2012). However, their findings may be limited to particular leaders or events and suffers a problem of inferring emotions from historical documents and personal recollections. Moreover, despite careful process tracing, they are unable to provide evidence for a causal effect of observed emotions on conflict behavior. More recent experimental studies have offered causal evidence for emotions and their mechanisms by isolating the variable of interest (e.g., Powers and Altman 2023; Renshon, Lee and Tingley 2017; Masterson 2022), but they entail concerns about generalizability as they are unable to establish direct links between emotions and events at the international level.

This brief discussion illustrates two outstanding issues in the emotions research in international relations. First, while the literature as a whole has adopted multiple methods, each individual study has tended to use a single methodological framework. Qualitative work and experiments each have their own strengths and weaknesses, but relying on a single method will subject a study itself to the limitations of the method it utilizes. Second, to my best knowledge, there is no quantitative cross-national analysis on emotions. This is in fact a data issue. Political scientists rarely have access to leaders' fleeting emotions in real time during crises, let alone at the cross-national scale (Koschut et al. 2017). This suggests that external validity should be a greater concern for scholars seeking to understand the role of emotions in international politics. When studying emotions, it is crucial to use multiple methods and data sources due to the significant heterogeneity in how individuals around the world understand, experience, and express emotions. Since emotions are greatly influenced by cultural and social norms, findings from one place cannot be assumed to travel to others. This also indicates a missed opportunity, as we have been unable to utilize existing data on interstate disputes and crises to understand how emotions broadly affect international conflict.

I address these challenges through a novel combination of a cross-national quantitative analysis of public statements and a survey experiment. By focusing on emotion *expressions*, I create an original dataset of world leaders' public statements during international crises between 1946 and 1996. The dataset includes hand-labeled anger expressions and threats in 5,807 statements recorded in the Foreign Broadcast Information Service (FBIS) and the Public Papers of the Presidents of the United States (Woolley and Peters 1999). This cross-national approach allows for testing different implications of emotional statements for threat credibility in a temporal and geographical scope that would not be feasible in a case study or an experiment alone.

The use of a randomized experiment allows me to identify the causal effects of anger expressions and examine causal mechanisms at the individual level. Previous case studies have noted that anger can communicate resolve (Wong 2019; Hall 2015). However, their observations are not conclusive as they do not empirically disentangle anger expressions from threats and material actions, making it difficult to identify the impact of anger expressions with precision. The nature of a historical case study also limits the generalizability of the findings. I address these gaps by using cross-national quantitative analysis and a survey experiment, providing both broad-based patterns and causal evidence.

1.5 Chapter Outlines

Chapter 2 develops a theory of anger expressions as a mechanism for the credibility of threats to use force. I first provide an overview of how the international relations scholarship characterizes threats as a signal of resolve in foreign policy crises, highlighting the limitations of costly signaling and the theoretical advantages of a theory of credibility based on emotions. Then I delve into how leaders' emotional expressions, specifically anger, function as supplementary information about resolve to make threats credible in the eyes of the target. Drawing on insights from social and evolutionary psychology as well as political psychology, I will lay out the socio-functional logic of anger and its influence on observers' beliefs and behavior. In doing so, I specify two key mechanisms through which anger expressions might affect beliefs about credibility. First, anger generates an image of strong resolve by making a leader appear relatively insensitive to the cost of conflict. Second, anger increases the reputational consequences of threats, making it more likely for threatening leaders to carry out their threats. Based on this theory, I derive six specific hypotheses that will be tested in three empirical studies

presented in Chapters 4 through 6.

Chapter 3 describes my original data on world leaders' public threats and anger statements. I have noted that one significant challenge in the empirical investigations of emotions in international relations is a lack of cross-national observational data. I collected the original data on public threats and anger statements to overcome this challenge. This chapter will define anger expressions and threats more clearly and discuss measurement strategies and data collection procedures. It describes in detail how threats and anger expressions are coded. I will provide descriptive statistics of the data, such as the distributions of anger statements, to highlight variation in the amount of anger leaders express in international crises.

Chapter 4 presents the results of the analysis using the data on public threats and anger statements. The statements data are combined with the data on international crises and political leaders to create a directed dyadic data set. Through a quantitative analysis of this data, I provide evidence that public threats embedded in greater levels of anger lead to higher chances of winning in international crises and conduct exploratory analyses to address confounding issues, such as reverse causation. This finding is robust to the inclusion of key covariates, including relative power, regime type, and the level of violence, and across several different model specifications. The findings in this chapter establish broad-based evidence that anger expressions play an important role in international crises. Despite my careful treatment of potential confounding issues, however, due to the limitations of correlational analysis using observational data, I am unable to make causal claims about anger and its mechanisms. Chapters 5 and 6 use survey experiments to complement this chapter.

Chapter 5 describes findings from a survey experiment to study anger expressions. Building upon the correlational evidence from the previous chapter, the experiment examines the causal relationship between expressions of anger and

beliefs about resolve. The experiment uses a hypothetical crisis situation between the United States and Russia and finds that an expression of anger from a Russian leader significantly increases U.S. respondents' estimates of Russia's resolve, as well as the Russian leader's willingness to accept casualties to win the crisis. Additionally, I find evidence that threats may backfire when they are devoid of anger—U.S. respondents' beliefs about Russia's resolve decreased as a result of receiving a threat from a calm, composed Russian leader. Lastly, the chapter also presents the results of a causal mediation analysis, showing that the causal effect of the Russian leader's anger operates through perceptions about the leader's cost insensitivity. Together, these findings provide strong support for my theory that anger is a mechanism for how threats are perceived by targets.

Chapter 6 describes another experimental study designed to investigate the reputational mechanism of anger expressions in the context of domestic reputation costs. Before presenting the experiment, this chapter discusses in detail the rationale for focusing on domestic politics and offers a refined hypothesis. The experiment uses a experimental design popularly used in the experimental literature on audience cost theory (Tomz 2007*a*; Kertzer and Brutger 2016), and studies whether anger expressions from a U.S. president systematically alters the U.S. public perceptions of the reputational consequences of the president's behavior. I specifically investigate and find evidence that anger increases perceived reputational costs of backing down from a threat and perceived reputational gains of following through on a threat. I also find that these effects of leader anger are positively moderated by the respondent's hawkishness such that the effects of anger are stronger for hawks than doves. However, I do not find strong evidence that leader anger expressions lead to more severe audience costs, suggesting the need for future research. Finally, Chapter 7 summarizes the findings and implications of the project.

Chapter 2

A Theory of Anger as a Source of Credibility

The previous chapter established that although explaining when threats are credible has been a central research question in International Relations, there remains significant variation not properly accounted for by existing theories. For instance, crisis diplomacy often involves repeated interactions between the same states, as crises and disputes tend to occur due to some underlying unresolved grievances. Theories about structural factors that remain relatively stable across crises cannot explain variation in threat credibility within such cases. Furthermore, the literature has yet to find conclusive empirical evidence for the effects of theoretically prominent factors such as military power and regime type, with some scholars maintaining that they have little predictive power (Downes and Sechser 2012; Snyder and Borghard 2011). This leaves a significant gap for scholars and policy makers seeking to understand the conditions for peaceful coercion.

To address this shortcoming, I propose a more micro- and elite-level explanation for threat credibility: elite anger expressions. In this chapter, I present the theoretical argument that political leaders can harness the strategic value of anger

to increase the credibility of threats they issue, thereby influencing the beliefs of their adversaries and the outcomes of international crises they are involved in. I lay out two specific ways in which expressions of anger can improve credibility. First, the emotions of a threatening leader influence how her threats are perceived by targets. When anger is revealed in threats, it lowers the targets' perceptions of the leader's sensitivity to the cost of fighting. These perceptions in turn increase the estimates of the threat's credibility. Second, leader anger increases the country's reputations at stake, altering the strategic calculus of making threats in ways that make threats credible. The anger emotion conveys an actor's cognitive and behavioral commitment to attack its target Frank (1988). Consequently, following through on a threat embedded in anger results in greater reputational benefits as the leader confirms his or her commitment, and backing down from it leads to greater reputational costs.

I begin by discussing the conditions under which threats are thought to operate as a credible signal of resolve and highlighting why threats might still fail to convey resolve despite the conditions specified by international relations theories. I then draw on the evolutionary approaches to emotions to develop a model of anger in world politics, elaborating on two ways in which anger expressions influence observers' beliefs about resolve. Lastly, I conclude by discussing how anger expressions themselves might gain credibility and specifying the scope conditions of my theory.

2.1 Threats as a Signal of Resolve under Pervasive Uncertainty

Threats in the context of international crises are statements issued by state leaders indicating their resolve. By threats, I am referring to statements that make specific

promises to use force as well as statements that indicate demands, refusals, and policy stances without such promises. This breadth is justified because a statement can be a signal of resolve as long as it establishes a commitment to a position and suggests a possibility of fighting over the issue at hand. As such, my definition of a threat is very similar to that of "statements of resolve," which "indicate that a country is committed to a position" (McManus 2017), and broader than the definition of a compellent threat, which requires both a coercive demand for a policy change and a threat of force (Sechser 2011).

Before discussing how anger servers as a source of credibility, we should first address the question of why anger might be needed to make threats credible. After all, the logic of costly signaling provides compelling intuition that resolved leaders can separate themselves from unresolved ones by incurring costs that would commit them to following through because the unresolved types would not be willing to pay the costs. However, this logic relies on assumptions about the leader's ability to generate sufficiently high cost and the foreign target's ability to anticipate the costs *a priori*. These are nontrivial assumptions that have direct implications for credible communications of intentions because the sender of a threat and its target each face significant hurdles in communicating resolve.

Since states have incentives to make exaggerated claims about their resolve (Fearon 1995), leaders issuing a threat must demonstrate that they are genuinely committed to following through on it. As discussed in the introductory chapter, most of the existing theories about credibility invoke the logic of costly signaling to suggest that threats communicate resolve more credibly when they incur costs to the threatening leader. Domestic audiences may punish leaders who threaten to use force and fail to follow through (Fearon 1994). Adversaries that receive threats may make war preparations that hurt the leader's change of winning (Trager 2010), and countries that back down from a threat will develop a reputation for bluffing

(Sartori 2005). Regardless of the source of the cost, the act of issuing a threat carries little cost in itself; it is the reactions from the domestic and international audiences to the threat that create potential future costs. Consequently, the costliness of a threat is ultimately shaped by some combination of the leader's domestic political and material institutions (Schultz 1998; Weeks 2008; McManus 2017) and the specific contexts of the crisis in which it is engaged (Trager 2010).

On the other hand, the target of a threat should overcome the problem of misperception that arises from psychological biases and cognitive limitations inherent in the human mind. Contrary to what costly signaling theories assume, a statement whose intended message is obvious to the leader may not be so in the eyes of the target (Snyder and Diesing 1977). Recent studies suggest that elites are not exempt from these biases and limitations (Kertzer 2022). Dispositions and past experiences systematically affect the ways in which adversaries' statements are interpreted, and emotions can distort rational judgment (Yarhi-Milo, Kertzer and Renshon 2018; Mercer 2010). Moreover, the hectic nature of crisis diplomacy is likely to impede decision makers' ability to discern informative signals from noise, exacerbating the misperception of signals (Katagiri and Min 2019).

To illustrate the decision problem faced by the target of a threat, consider an assumption of the audience cost model that "leaders are able to generate arbitrarily large audience costs and so are able to tie their hands, no matter how great the expected costs of a military conflict," which Fearon Fearon (1997) acknowledges is "surely too strong." What is important for our purposes is that a threat commits a leader to fighting if the cost of backing down is larger than the cost of conflict. The target of a threat therefore faces a difficult task of determining the size of audience costs in case the threatening leader backs down and then inferring resolve from the estimated audience costs, while accounting for other situational factors that could go into the leader's costs-benefit calculations about the overall value of fighting.

Importantly, many of the factors, such as the leader's beliefs about the stakes or domestic support for using force, are not directly observable or only imperfectly known when a threat is issued.

In fact, estimating audience costs itself is likely to be challenging. While the classic audience cost theory suggests that audience costs arise due to the public's aversion to inconsistency, many recent studies have suggested that anticipating an adversary's audience costs requires knowledge about the characteristics of the audience. Experimental studies find that the public cares more about policy outcomes than inconsistency in their evaluation of presidential performance in foreign affairs (Nomikos and Sambanis 2019; Chaudoin 2014). Kertzer and Brutger (2016) show that the penalty for backing down significantly varies depending on such factors as the audience's foreign policy attitudes and social identities. After all, it is not unusual for leaders to miscalculate the level of support for war and foreign policy preferences among their own population—it would be unrealistic to expect foreign adversaries to fare better.

The discussion thus far has focused on audience cost theory, but my contention that it is difficult to assess costs may generalize to other factors that require some understanding of domestic political processes. This logic hints at the need to examine the possibility that targets of a threat often rely on more micro-level indicators to assess its credibility. Diplomacy scholars have put forth affect as a promising alternative explanation, providing archival evidence that political elites use their personal impressions of other leaders as a credible source of information about intentions (Hall and Yarhi-Milo 2012; Yarhi-Milo 2013). Building on this research, I propose expressions of anger as a previously underappreciated determinant of threat credibility. Studying a specific emotion like anger provides an advantage over generalized affect in that the behavioral consequences of the emotion are more clearly defined, allowing for the formulation of clearer hypotheses.

2.2 Anger as a Commitment Device

2.2.1 Anger expressions

Anger is a negative feeling that arises in response to perceived harm by another agent to an important personal or social value (Berkowitz and Harmon-Jones 2004). Anger is an intense, context-specific feeling that has an identifiable trigger. Anger also has a significant moral dimension: Individuals feel anger not just when they perceive harm but when they believe that the harm is deliberate, unfair, and unjustified. Another important dimension of anger is a sense of control. Individuals are more likely to feel anger, rather than sadness or fear, when they believe they can deal with the triggering agent (Lazarus 1991; Frijda 1986). Thus, as appraisal theories suggest, the situation itself does not produce anger in a predefined way. Rather, individuals' cognitive assessments of the situation's implications for their values determine the extent to which anger is elicited.[1]

Anger can be expressed through physiological signals as well as verbal communications. As Wong (2019) puts it, observers detect anger from "a repertoire of behavioral cues expressed on an actor's face (furrowed eyebrows, pressed lips, flared nostrils, flushed face, glaring eyes, etc.), and through his voice (raised and rising pitches, growling, etc.) and body (clenched fists, foot-stomping, intense trembling, etc.)." An extensive body of research shows that humans are naturally capable of recognizing emotions in others through such behavioral cues with little cognitive effort (Tracy and Robins 2008; Aviezer et al. 2011). In addition, many of the experiments on emotion recognition successfully used video clips

[1]Additionally, anger is believed to be a "basic" human emotion (Ekman 1992*a*; Plutchik 2001). Basic emotions are innate and universal, with distinct physiological characteristics. Emotion theorists disagree on what basic emotions there are, but most include anger. Ekman (1992*a*) proposes six basic emotions: anger, sadness, joy, fear, disgust, and surprise. Plutchik (2001) adds two more emotions to Ekman's list, trust and anticipation. This point is empirically useful. Studying a basic emotion allows me to sidestep the issue of cultural and semantic differences in emotions.

or photos of anger displays as stimuli in lab environments. This fact implies that people are able to understand others' emotions through impersonal channels in which leaders' statements might be delivered, such as televised speeches and radio broadcasts.

In addition to nonverbal expressions, leaders rely on verbal communications to clarify the cause, meaning, and target of the anger. Leaders might be explicit in their expressions, as President Carter said in 1980 about the Iranian hostage crisis that "We are outraged when we see our own citizens kidnaped and abused in Iran" (January 11, 1980, quoted in Woolley and Peters (1999)). But anger is often conveyed indirectly and manifests itself in a discourse of grievances, accusations, or insults (Lerner and Tiedens 2006; Quigley and Tedeschi 1996). When asked to comment on the hostage crisis again in a subsequent interview, Carter replied, "It's a criminal act: a group of terrorists, kidnapers, seizing innocent victims and holding them for attempted blackmail in an unprecedented way, supported and encouraged by government officials themselves" (January 20, 1980, quoted in Woolley and Peters (1999)). Criticizing the international community's failure to stop the the Assad regime from killing civilians during the Syrian civil war, Turkey's president Erdogan reportedly said, "I wonder how long you will turn a blind eye to this massacre. ... Damn your international policies!"[2] These example show that verbal communications of anger are distinct from threats and related concepts like "saber rattling," which represents a broad category of hostile rhetoric (Wood 2012). Leaders expressing anger do not always antagonize the target, make coercive demands, or threaten to use force. Leaders are also capable of issuing threats unemotionally and lashing out without making specific threats.

[2]https://www.aljazeera.com/news/2013/5/7/erdogan-israeli-raids-aided-syria-cover-up

2.2.2 The functional logic of anger

Anger is thought to have evolved to prepare humans to deal with social and material problems they are faced with by inducing behavioral states that are suited for the circumstance, often by overriding the current behavioral states and pre-existing dispositions (Frijda 1986; Lazarus 1991; Frijda and Mesquita 1994). Originally put forth by (Frank 1988), this evolutionary model of anger explains that because conflict is costly, purely rational individuals are unable to credibly commit to punishing transgressors. However, if individuals act in anger and disregard their immediate welfare, others will be less likely to treat them in a way that they believe will trigger anger. Consequently, angry individuals may receive better treatment in the long run. To achieve this, anger evokes cognitive and physiological responses geared toward aggression, inducing hostile, risk-seeking attitudes (Berkowitz 1990; Lerner and Keltner 2001) and willingness to bear costs to punish the blameworthy actor (Harmon-Jones and Allen 1998; Carlsmith, Darley and Robinson 2002). Similarly, Hutcherson and Gross (2011) argue that anger prompts individuals to respond to threats with "more proactive, but potentially costly, strategies." Analyses of ultimatum games have shown that angry bargainers often reject unfair offers when accepting them would produce better material outcomes (Pillutla and Murnighan 1996; Sanfey et al. 2003). This behavior is also observed in a survey experiment in the context of international crises and even when the bargainer's reputation is not at stake, suggesting the strong commitment effect of anger regardless of the consequence of the behavior (Yamagishi et al. 2009; Powers and Altman 2023).

If *experiences* of anger serve as a commitment device, then *expressions* of anger should be an effective bargaining tool. According to the recalibrational theory of anger, anger signals that one has been treated unfairly and threatens to impose costs if the unfair treatment continues (Sell 2011; Hall 2015). And targets of anger

tend to view this threat as credible: Research has shown that people intuitively perceive others' anger displays as a predictor of destructive behavior (Van Leeuwen et al. 2018). In a similar vein, research on negotiation suggests that people use another person's anger expression to make inferences about the lowest price that person would accept in bargaining (Hareli and Hess 2010; Keltner and Haidt 1999; Van Doorn, Heerdink and Van Kleef 2012). Equally important, because the human capacity for anger is derived from evolution, this socio-functional logic of anger appears to be understood by people across cultures (Sznycer, Sell and Dumont 2022; Alonso-Arbiol et al. 2011).

A key premise to my theory is that expressions of emotions provide social information. A recent literature in social psychology suggests that people use an observed emotion to understand the person's underlying cognitive processes (Van Kleef, De Dreu and Manstead 2010; Hareli and Hess 2010; Van Doorn, Heerdink and Van Kleef 2012). This "reverse appraisal" theory is built on appraisal theories of emotions, which posit that a person's emotional experiences arise from her evaluation of the situation in relation to her goals, desires, and beliefs (Monroy, Cowen and Keltner 2022) (e.g., Is this situation relevant to and congruent with my goals? Who caused this situation? Am I able to resolve this situation?). Appraisal theories suggest that an event does not map to specific emotions in a predetermined way. Instead, a person's subjective appraisal of the situation determines the type and intensity of emotions that will be experienced and, consequently, the behavioral response to the situation. Therefore, as Van Kleef, De Dreu and Manstead (2010) argue, if "each discrete emotion has its own antecedents, appraisal components, relational themes, and action tendencies," then "observing a particular emotion in another person provides relatively differentiated information about how that person regards the situation." For example, individuals may feel frustrated in a challenging situation if they believe they are

failing to realize their goal. Those who believe that the situation is irrelevant to their goals and those who view the situation as manageable will not feel the same way. A display of frustration reveals information about individuals' valuation of the issue and beliefs about their ability to carry out the task.

Research on negotiations provides experimental findings supporting the reverse appraisal theories. Overall, the findings show that people extract information about cognitive appraisals from others' emotions and utilize this information to make inferences about their beliefs, preferences, and intentions (Hareli and Hess 2010). De Melo et al. (2014) show that observers use expressed emotions of others to infer their cooperative intentions, with this inference being mediated by observers' beliefs about others' appraisals of the situation. Van Doorn, Heerdink and Van Kleef (2012) find that expressions of anger, compared to happiness or disappointment, lead observers to think that the expresser is interpreting the situation as more competitive. In addition, people not only draw inferences about others' intentions but make strategic decisions based on that information. A series of bargaining experiments show that bargainers make more concessions to angry opponents, and that bargainers adjust their demands based on their opponents' "bottom line" inferred from their emotion expressions (Lelieveld et al. 2012; Van Kleef, De Dreu and Manstead 2004; Lelieveld et al. 2011).

At face value, actors in world politics appear to process emotion information in a manner that is consistent with the reverse appraisal theory. Scholars of diplomacy have documented a number of historical cases in which leaders, diplomats, and state officials use emotional and affective cues as signals of intentions of one another (Holmes 2013; Hall and Yarhi-Milo 2012). This body of literature suggests that leaders often form judgments about their opponents' intentions based on personal impressions, and that expressions of anger by leaders can affect other leaders' beliefs regarding resolve, even in high-stakes situations such as the Berlin

Crisis (Hall and Yarhi-Milo 2012; Wong 2019). Leaders indeed seem to behave as if others can read their intentions from expressed emotions, even beyond face-to-face interactions. In September of 1994, for example, President Bill Clinton made an angry speech directed at Haiti's military dictator General Cedras, calling him a leader of "armed thugs" and accusing him of gross human rights violations while threatening to forcibly remove him from power. An unnamed administrative official commented on Clinton's anger, saying, "We want them [the Haitian military leaders] to watch the president themselves, watch what the president's words are—and how he says them. Have them see that this is serious."[3]

Transferred to the domain of international crises, the discussion thus far suggests that anger may be a key mechanism for how threats are perceived by their targets. If resolve represents a willingness to bear costs in pursuit of a policy (Ramsay 2017), an angry leader may project an image of resolve by appearing insensitive to the cost of conflict. However, anger, being an internal emotional state, is not directly observable. By embedding threats within expressions of anger, leaders may ensure they project that image. Thus, anger expressions will increase a threat's credibility despite uncertainty about the cost of backing down. This argument is consistent with recent research showing that leaders make more credible public threats when they are militarily stronger and more secure in office because they incur fewer material and political costs for entering a conflict (McManus 2017).

Although expressing anger is unlikely to affect the material costs of fighting, it allows the adversary to tap into the leader's mental states and make inferences about how she might appraise the costs (De Melo et al. 2014; Keltner and Haidt 1999). Anger may add credibility to a threat even when following through on it appears to be against the leader's best interest. At the onset of the Libyan

[3]https://www.baltimoresun.com/news/bs-xpm-1994-09-15-1994258005-story.html]

civil war in 2011, Libyan leader Moammar Gadhafi furiously threatened that he would rather "die here as a martyr" than give up power (Meikle and Black 2011). Immediately after, Western observers characterized Gadhafi as "a dangerous nut," who could "turn himself into a suicide bomb and take his country down with him" (Memmott 2011). Conversely, appearing cost sensitive undermines credibility even for leaders with military superiority. Before the Iraq War in 2003, for instance, Saddam Hussein doubted Bush's resolve "because of exaggerated US fears of casualties" although Hussein had "no illusions about US military or technological capabilities" (Duelfer 2005).

The theoretical discussion thus far yields specific observable implications about anger expressions both at the international level and the individual level. Crisis bargaining is often characterized as "contests of resolve," in which each side will compare its own resolve to the opponent's perceived resolve when deciding to back down or continue the crisis (Morrow 1989). In the end, the side that demonstrates greater resolve convinces the other to concede. If my theory is correct, we would expect to see evidence of increased credibility reflected in the outcome of the international crisis in which the threats are issued. An adversary that receives a threat from an angry leader will be more likely to believe that the leader is resolved to fight. As the adversary revises upward its beliefs about the leader's willingness to stand firm, the adversary should become more likely to back down in the crisis unless it is willing to escalate. Therefore, holding everything else equal, I expect that threats embedded in anger expressions will lead to a higher chance of winning in a crisis.

Hypothesis 1: *Threats accompanied by higher levels of anger expressions will be associated with higher chances of prevailing in international crises.*

Considering the psychological challenges of coercion (Dafoe, Hatz and Zhang 2021; Powers and Altman 2023; Hall 2017), it is possible that expressing anger

causes the threat to backfire by provoking counter-anger in the target. However, as the experiment in Chapter 5 will demonstrate, people experience not only anger but also fear and anxiety in response to an adversary's anger, the emotions that are known to increase risk-averse and conciliatory attitudes (Lerner and Keltner 2001; Van Kleef, De Dreu and Manstead 2004). Moreover, the current literature on emotions does not provide clear theoretical expectations or empirical findings about the combined influences of emotions on decision making (Ferrer and Ellis 2019). On the other hand, focusing on emotional arousal rather than specific emotions, Renshon, Lee and Tingley (2017) find that bargainers who are aroused by the opponent's unfair offer show limited abilities to reason strategically, making sub-optimal decisions that result in lower payoffs but prevent conflict. Thus, without clear guidance on the effect of the complex interplay of emotions, we are on firmer ground in hypothesizing about the informational effects of anger expressions than we are in making predictions about the emotional responses of targets.

In addition to the hypothesis focused on crisis outcomes, we expect more direct evidence at the level of individuals' perceptions and beliefs. I thus expect that expressions of anger will increase beliefs about resolve. Moreover, if anger is a mechanism for how threats are perceived, threats delivered in anger will lead targets to believe that the threatening adversary has greater resolve, compared to the same threats without anger. We could also hypothesize about the credibility-enhancing effect of anger from the angle of uncertainty. Within the framework of international crises as contests of resolve, uncertainty and incomplete information about the opponent's resolve cause crises and escalation into war (Morrow 1989), and a credible threat that convinces an adversary to back down should do so by reducing the adversary's uncertainty about the threatening leader's resolve. Therefore, I expect that targets that receive threats accompanied by anger will

become more certain of the adversary's resolve.

Hypothesis 2: *Expressions of anger will increase beliefs about resolve.*

Hypothesis 3: *Targets who receive threats accompanied by anger expressions will become more certain about their estimates of resolve than those who receive the same threats not accompanied by anger.*

Lastly, I derive a hypothesis about one mechanism for how anger expressions operate. I have proposed that leaders displaying anger project an image of resolve by appearing relatively insensitive to the cost of conflict. I will test this argument through a survey experiment using causal mediation analysis (Imai and Yamamoto 2013), which investigates the extent to which the causal effect of a treatment on the outcome flows through an intermediate pathway. To this end, I hypothesize that the perception that the leader is willing to accept the cost of conflict will mediate the effect of anger expression on beliefs about resolve.

Hypothesis 4: *Perceptions of willingness to accept the cost of conflict will mediate the effect of anger expressions on estimates of resolve.*

2.3 Leader Anger and Concern about Reputation

One significant advantage of anger over other sources of credibility is that it is intuitive. Unlike costly signals, anger conveys resolve through humans' innate understanding of the evolutionary function of the emotion. While we might characterize anger expressions as "cheap talk," both the expressions and the perceptions of anger are, to a large extent, not always under conscious control. This automatic property of anger enables the communication of private information between adversaries. In international crises, faced with a computationally intractable

task of assessing resolve (Kertzer, Renshon and Yarhi-Milo 2021), observers may rely on their adversary's emotions as an intuitive indicator of intentions.

Yet, even if anger is readily manipulable, and even when targets can accurately assess the cost of making threats, anger expressions may still have meaningful impact because they can alter the strategic consequences of threats. Specifically, anger expressions increase the reputational considerations at stake. This makes empty threats more costly and genuine threats more rewarding, incentivizing the threatening leader to follow through. To understand the relationship between anger expressions and reputation, we first need to discuss how reputation arises and how it influences the conduct of international crises.

2.3.1 What is reputation?

Reputation is "a belief about a trait or behavioral tendency of an actor, based on that actor's past behavior" (Dafoe, Renshon and Huth 2014). While reputation can arise with regard to any trait, I focus specifically on what Jervis (2017) calls "signaling reputation"—a leader's reputation for "doing what she says she will do," such as fulfilling threats and promises. Making a threat and then failing to follow through on it may lead others to infer that the leader considers the issue in question not very valuable after all, and that future threats she makes in similar situations would likely to be empty. On the other hand, leaders with a positive signaling reputation will be believed to be more likely to follow through on their promises to use force if their demands are not met by adversaries. Also known as "reputation for honesty" (Sartori 2005) or "reputation for action" (Schelling 1966), a signaling reputation is believed to be a powerful bargaining tactic that "enables the possessor to construct new commitments through a simple statement" (Dafoe, Renshon and Huth 2014). Indeed, the concern about reputation features prominently in existing theories about threat credibility: leaders increase the cost

of backing down by staking their reputation on carrying out threats (Schelling 1966).

Because reputation has long been considered a major motive for war, scholars of international relations have fiercely debated the degree to which reputations affect decision makers' behavior in international conflict, as well as the very existence of reputations. (Mercer 1996; Press 2005; Huth 1997; Weisiger and Yarhi-Milo 2015; Tomz 2007*b*; Harvey and Mitton 2016). For our purposes, however, two issues are particularly relevant. First, while the international relations scholarship does not agree on whether reputation is specific to states (Sartori 2005; Tomz 2007*b*; Walter 2006) or leaders (Lupton 2020; Guisinger and Smith 2002; Yarhi-Milo 2018; Wolford 2007), I discuss reputation in terms of states without taking a position in the debate. According to Renshon, Dafoe and Huth (2018), observers tend to assign reputation both to leaders and governments based on past action, although reputations adhere more strongly to the actor that is more influential at the decision making task at hand. Brutger and Kertzer (2018) also find that U.S. citizens form beliefs both about the country and the leader in response to the leader's crisis behavior. Therefore, although a state cannot feel anger, there is evidence that reputational considerations stemming from the leader's anger can adhere to the state.

Second, theories inspired by political psychology provide insights on the critical role of expectations in the logic of reputation. Whether a leader gains or loses a reputation depends on others' pre-existing beliefs about her commitment (Harvey and Mitton 2016; Mercer 1996). As Dafoe, Renshon and Huth (2014) put it, "[i]f no one perceives the US security commitment to Japan to include defending Japan's claims to the Senkaku/Diaoyu islands, then failing to do so will cause little harm to the United States' reputation." In addition, past action has greater impact on beliefs (reputation) when the action runs counter to others' expectations (Tomz

2007*b*; Snyder 1961; Huth 1997). This is because observers factor in available information about a leader, such as interests and capabilities, when forming expectations about the leader. If, for example, an adversary expects a threatening leader to use force but the leader fails to follow through, this may lead the adversary to infer that the leader lacks resolve for reasons that are not accounted for by the information available to the adversary, resulting in a downward revision of beliefs about resolve. In contrast, if the adversary expected the threatening leader to be bluffing in the first place, failing to carry out the threat would merely validate the adversary's pre-existing beliefs. The leader would still lose reputation as long as there was a non-zero chance of the threat being sincere, but the leader's irresolute behavior would have smaller impact on beliefs about resolve in this case.

2.3.2 How anger increases concern about reputation

Leaders have strategic incentives to live up to their threats in order to build and preserve signaling reputations. Expressions of anger allow leaders to leverage the reputations to enhance their coercive credibility in two ways. First, anger raises observers' expectations about a leader's commitment to fight. This is a direct implication of the previous discussion about emotions as an evolutionarily derived commitment mechanism. Intuitively, if leaders lose reputation when they fail to meet others' expectations, then a perceived stronger expectation of force will lead to greater reputational consequences. This claim is consistent with how reputations are treated in the literature on audience costs, which are the punishment leaders suffer when they renege on policy commitments that they have publicly made (Tomz 2007*a*). While there are multiple mechanisms underlying audience costs (Nomikos and Sambanis 2019; Kertzer and Brutger 2016), the public's concern about their state's reputation is regarded as the most prominent motive for punishment in many traditional models of audience costs (Fearon

1994). The reputational concern in turn is caused by the leader's inconsistency between her threats and actions. Importantly, although whether to carry out a threat or not might be a binary decision, perceptions of inconsistency likely exist on a continuous spectrum and can vary depending on the discrepancy between words and actions (Lin-Greenberg 2019). Fearon (1997) implies that threatening leaders may make "full" or "halfhearted" commitments by using either ambiguous or specific language. Experimental findings also suggest that more specific threats induce greater audience costs (Trager and Vavreck 2011). When the leader expresses anger and then backs down from his threat, this creates greater inconsistency between her words and others' expectations, hurting the country's reputation to a greater extent. On the other hand, when a leader acts on anger in following through on his threat, this validates others' expectations, allowing the leader to develop a reputation for resolve.

Second, anger may add another layer of reputational loss when the leader fails to carry out the threat. If anger behaviorally prepares the leader for aggression, then choosing not to fight may indicate that the leader has feigned or exaggerated anger. In effect, the leader may gain a reputation for being emotionally dishonest, which would hurt the leader's emotional credibility in the long run. Even if the sincerity of the emotion is not doubted, it would still suggest to others that the leader is capable of suppressing the influence of anger in decision making. In either case, the strategic value of anger expressions will diminish, as adversaries in the future may believe that anger is unlikely to affect the leader's decisions about using force. Furthermore, the effect of a reputation for emotional dishonesty might spill over to other traits and issues, hurting the general credibility of the leader in the eyes of friends and foes alike. Multiple studies have shown that people who exaggerate emotions is generally viewed as untrustworthy, manipulative, and unethical (Cheshin, Amit and Van Kleef 2018; Szczurek, Monin and Gross

2012). Insincere emotion displays damage trust between cooperative partners (Côté, Hideg and Van Kleef 2013; Van Beest, Van Kleef and Van Dijk 2008) and make opponents more intransigent in bargaining (Hideg and van Kleef 2017; Tng and Au 2014; Campagna et al. 2016).

In contrast, fulfilling a threat in anger might grant the leader a reputation for being easily angered. This reputation can be characterized as general or specific based on what inference observers make from the expressed anger. When observers attribute the leader's behavior in anger to her disposition, the leader may cultivate a reputation for being a "hothead," or, in psychological terms, high in trait anger. Trait anger is an individual's "chronic tendency to experience the emotion of state anger with greater frequency, intensity, and duration" (Veenstra, Bushman and Koole 2018). People with high trait anger tend to view situations around them as threatening. They are less capable of controlling their hostility and hence possess heightened motivation to attack what they see as threatening agents. Since trait anger is an individual disposition, this reputation concerns others' general perceptions about the leader's general "type," analogous to reputations for "resolve," "toughness," and "reliability," rather than specific beliefs about the likelihood of following through on threats.

In international relations, a reputation for trait anger may not only increase the perceived authenticity of the leader's emotional expressions but also serve as a general reputation for resolve that is transferable between crises. Regardless of whether the leader has previously issued threats or displayed anger, she is likely to be viewed as capable of lashing out violently, possibly even in a self-destructive manner. This perception may foster fear of a reckless use of force with little regard for costs, convincing potential adversaries to moderate their demands in disputes to avoid triggering anger.[4]

[4]Wong (2019) provides a counterargument to this argument. He argues that being a hothead carries little strategic benefit because chronic anger is a noisy signal that does not distinguish

Specific reputations for anger may emerge when observers explain the leader's anger using situational attributions (Mercer 1996). A leader may act on anger in crises she is involved but display an even-tempered disposition in other diplomatic interactions. It could also be that leaders communicate more intense anger in crises that implicate core national interests than in crises that are less salient. Since crises are uncommon, the leader will express anger relatively rarely. In this case, observers may attribute the anger to high interests at stake in the crisis rather than to the leader's individual dispositions. Therefore, specific reputations for anger operate through re-shaping adversaries' estimates of the leader's interests. If people experience anger when they are prevented from achieving their goals (Berkowitz and Harmon-Jones 2004), then anger conveys information about the leader's valuation of the issue in dispute. The evolutionary theory of anger suggests that the intensity of anger is roughly proportional to the perceived harm (Sell et al. 2017). The more emotionally invested a leader is in the issue at stake, the angrier she will become in response to a perceived breach of it (Ryan 2017; Tetlock 2003). Although the reputation developed this way might only be relevant to crises involve similar issues, it can still impact adversary's beliefs about credibility. After all, as Weisiger and Yarhi-Milo (2015) argue, reputations are likely already "folded into the general assessments of interests, alongside other pertinent sources of information."

In sum, leaders may strategically put their reputation at greater stake by displaying anger when making threats to use force. Anger raises expectations about fulfilling those threats, increasing the inconsistency and hence the resulting

---

between the blameworthy and the innocent. He also argues that anger expressions from a hothead may backfire as others would view the emotion as insincere. This argument is relevant if we see anger primarily as a signal. However, from a reputational perspective, if an actor is known to be quick to anger, then there is no logical reason to expect that the informational value of the reputation will diminish. We can think of a reputation for anger as being intrinsically aggressive, or having a low baseline cost of conflict. Moreover, if a leader expresses anger and then behaves belligerently, the perceived authenticity of the emotion will in fact increase.

reputational concern if the leaders back down. Additionally, I have proposed the concept of a reputation for anger, a type of signaling reputation that could be utilized to provide bargaining advantages in crises. Of course, the concept will be useful only if anger provides strategic benefits as my theory predicts, so my hypotheses about this reputation-based mechanism of anger will likely depend on if we can find evidence for the first set of hypotheses.

This argument suggests that leader anger amplifies the reputational consequences of making threats—as a result, a leader who threatens force in anger has a greater incentive to see through the threat, and their adversary is more likely to believe their resolve. Since the effect of anger on beliefs about resolve will be tested via the preceding hypotheses, I focus on reputational consequences. I hypothesize that an angry threat incurs greater damage on the country's reputation when the leader fails to carry out the threat and strengthens the country's reputation when the leader fulfills the threat.

Hypothesis 5: *Threats accompanied by anger expressions will result in greater reputational consequences, compared to the same threats without anger.*

2.4 How anger expressions gain credibility

How might leader anger gain credibility when expressing it could produce strategic benefits? Generally speaking, emotional expressions are neurological responses that are difficult to feign or inhibit voluntarily (Darwin 1872; Ekman 1992*b*). Although most individuals, including political leaders, may regulate their emotions to a certain extent following social norms (Ekman and Friesen 1971), physiological cues, such as sudden body movements, blushing, gasps, raised eyebrows, and changing tone of voice, are not always under conscious control and reveal a genuine experienced emotion or lack thereof (Ekman and Friesen 1974). Ver-

bal mistakes within a speech or mismatches between words and nonverbal cues can also betray emotions. People's inability to manipulate emotional displays is so common that some consider it "essentially ubiquitous," especially when the emotion is intense (Porter, Ten Brinke and Wallace 2012).

The involuntary nature of emotional expressions makes them what Jervis (1970) refers to as *indices*, behaviors that carry inherent credibility by virtue of being largely non-manipulable. Emotional indices are still subject to error because individuals vary in their ability to communicate and recognize emotions. Nonetheless, leaders rely on other leaders' affective cues as credible indicators of intentions to a significant extent (Holmes 2013; Yarhi-Milo 2013; Hall and Yarhi-Milo 2012). Leaders may indeed have unusual confidence in their ability to discern others' emotions, attributing their rise to power to their keen judgment about others (Jervis 1970). After meeting with Vladimir Putin in 2001, George Bush said that he "looked the man in the eye" and "found him to be very straightforward" (Mufson 2015). After reading Hitler's face as he stated that he would not invade Czechoslovakia, Neville Chamberlain described him as "a man who could be relied upon when he had given his word" (see Ekman 1992*b*). However inaccurate their assessments may have been, both Bush and Chamberlain appeared confident in their interpretations of their counterparts' emotions.

2.5 Scope Conditions

Expressions of anger are most useful for threatening leaders under certain conditions. First, anger is more effective when threats are less credible. Any increase in credibility that results from an anger expression would be marginal if a threat is already believed. Bargaining studies have found that angry expressions increase the credibility of threats and complaints and thus lead to better offers and com-

pensations, but only when the threats and complaints themselves lack credibility (Reed, DeScioli and Pinker 2014; Hareli et al. 2009).

Second, the target must not know that the leader is angry. Emotional expressions would convey little new information when the actor's emotional state is already known. Negotiation research indicates that anger expressions were less effective at drawing concessions when they come from bargainers already known to be angry (Filipowicz, Barsade and Melwani 2011), a pattern observed among political elites in crisis bargaining (Wong 2019). In addition, anger may have higher informational value when communicated to unsuspecting targets because observers tend to learn more from an actor's behavior when it is surprising (Burgoon 1993).

Consequently, anger expressions may be useful in most cases for challengers, but only in some cases for defenders. Defenders might generally make credible threats without the help of anger due to various psychological and materials reasons. For example, if leaders are risk-accepting to avoid losses but risk-adverse to obtain gains, as prospect theory suggests (McDermott 2001), then threats to defend the status quo should be more credible than threats to overturn it. The burden of projecting power overseas may also weaken the credibility of challengers' threats (Schelling 1966). Furthermore, leaders intent on challenging other leaders may well anticipate anger, learning little from their targets' angry reactions.

Finally, I expect that the logic of anger applies to most types of crisis. Just as leaders can engage in international disputes over various policy issues and threaten to take a range of actions, anger can accompany almost any coercive threat to increase its credibility, provided that following through on it is costly. Militarized crises may involve particularly high costs, but non-militarized crises often entail significant economic and political consequences, such as imposing sanctions and severing diplomatic relations. Study 1 uses data that include both

militarized and non-militarized crises to test my theory.

2.6 Conclusion

This chapter has introduced my theoretical argument that leader anger embedded in threats is a key determinant of the credibility of threats. Subsequent chapters will test this theory using a combination of cross-national analysis of original data on world leaders' public statements and survey experiments. Chapter 3 will discuss the data collection procedure and provide basic descriptions of the data. Chapter 4 will use the data to test Hypothesis 1 in order to provide correlational evidence for the role of anger in international crises. Chapter 5 will use a survey experiment to test Hypotheses 2 through 5 to provide causal evidence for the effects of anger and its mechanisms. Chapter 6 will employ another survey experiment to test Hypothesis 6 to further examine the mechanism of anger expression.

Chapter 3

Data on Public Threats and Anger Statements

Systematic empirical investigations of emotions in international politics face significant challenges in no small part due to lack of observational data. Qualitative case studies, which have been the dominant method in the literature, can shed light on the impact of emotions on foreign policy by tracing leaders' decision making processes (e.g., Wong 2019; Hall and Yarhi-Milo 2012). However, findings from case studies tend to be specific to particular leaders or historical events. The use of archival research also suffers the problems of inferring emotions from potentially biased historical records or recollections of leaders (Renshon, Lee and Tingley 2017). Experimental methods have recently gained popularity due to their advantage in identifying the causal effects of emotions, but they are unable to speak directly to actual international outcomes. Moreover, similar to case studies, experimental methods raise concerns about external validity as many international relations experiments rely on public samples recruited in a single country, typically in the United States. The external validity concerns are especially problematic in emotions research due to emotions' social nature—the norms of expressions

and meanings of emotions might vary considerably across time and space. Thus, existing scholarship has not been able to provide a comprehensive picture of whether and how emotions influence international outcomes.

Understandably, it is rarely feasible for political scientists to access political leaders' emotions in real-time during any international event (Koschut et al. 2017). One methodological innovation of this project is that it focuses on *expressed* emotions from leaders rather than *experienced* emotions.[1] This approach offers significant measurement advantages because emotional displays are much more observable than internal emotional experiences. This allows me to develop a cross-national measure of elite usage of anger using publicly available text data on leaders' political statements. Finally, focusing on elite communications of anger allows me to avoid taking a stance on thorny unit-of-analysis issues such as whether emotions can be imputed to states (Sasley 2011; Mercer 2014) and how individual emotions aggregate to impact collective decision making (Hafner-Burton et al. 2017).

3.1 Defining Anger Statements

As discussed in the previous chapter, anger can be communicated through bodily, physiological, or verbal expressions. This book focuses on verbal expressions, or what I call "anger statements." Anger statements are public statements indicating that the leader and/or the collective entity the leader represents (e.g., the public) is experiencing anger due to a perceived transgression by the adversary. I focus on verbal communication because it is logistically easier to obtain historical statements in text form than visuals or audio. As such, my primary data source, which will be described in subsequent sections, contains transcripts of radio broadcast

[1] For similar conceptual approaches, see Wong (2016) and Hall (2015).

or televised statements from 62 countries during the period between 1946 and 1996. That said, the transcribers and translators of the statements had access to the actual audio or visual materials real-time, and they made meticulous efforts to capture the tone and nuance of the original statements (Leetaru 2010).

I define anger statements in terms of public statements because of the empirical scope of this study, but leaders can express anger through private diplomatic interactions. My arguments about the informational effect of anger expressions apply to either public or private statements because they do not depend on audience costs. Scholars of diplomacy suggest that in face-to-face diplomacy, leaders use other leaders' affective cues as a credible source of information about their intentions (Holmes 2013; Wong 2019). In this framework, emotions are considered an involuntary information-revelation device since humans are unable to fully control their emotions (Wong 2016). In fact, if leaders can more easily communicate emotions in face-to-face interactions than public channels, we would expect that anger expressions will play a larger role in private diplomacy. However, since the empirical focus of this book is public statements, extrapolating the findings into private diplomacy should only be done with caution.

I focus on public statements because they are much more accessible than private diplomatic records. This easier accessibility provides a methodological advantage rather than simple logistical convenience. Given the relative dearth of private statements and the costs associated with collecting them, including private statements would significantly limit the spatial and temporal scope of my data. Although using declassified documents will shed light on elite perceptions of international signals unadulterated by public pressure (Katagiri and Min 2019; Schub 2022), such an approach typically focuses on within-case variations. Collecting declassified materials on the cross-national scale will be extremely challenging, if not impossible. As I established earlier, since the emotions research in international

Statement category	Examples
Direct expressions of anger	"**We are outraged** that, half a world away, *the Iranian Government holds 50 innocent Americans hostage*." Jimmy Carter, USA vs Iran, January 13, 1980 "It is not only our people *who will not forget it* and **who are indignant at it** but I already know that abroad also there is indignation at such an action." Josip Tito, Yugoslavia vs Italy, October 11, 1953
Indirect expressions of anger	"**We are losing our patience**. *They* [Costa Rican guerillas] *have kidnaped and taken many of our people and have not returned them to us*." Anastasio Somoza García, Nicaragua vs Costa Rica, December 29, 1978 "**He** [Tunku Rahman] **said that Indonesia is a madman and that Sukarno is the leader of madman. . . . We are called madmen, madmen, madmen. . . . Oh my God, even Leimena is called a madman. We are called madmen**." Sukarno, Indonesia vs Malaysia, March 14, 1964.
Verbal assaults	"**[T]his queer and mad man** [Gaddafi] does not know what it means to play with the armed forces. *How could he allow his forces, which were supposed to guard our unarmed forces, to arrest them* [Egyptian air force cadet]?" Anwar Sadat, Egypt vs Libya, July 22, 1977 "And I think we had a marvelous Christmas present when we got the word that Noriega [the Panamanian dictator], **the drug trafficker**, was in the Nuncio. And I am determined to bring him to justice." George H. W. Bush, USA vs Panama, December 27, 1989
Accusations of wrongdoing	"This is the Peruvian truth, the truth with which **it tries to deceive the world. . . . They are ready for a cease-fire while they continue attacking Ecuadorean territory.**" Jaime Roldos, Ecuador vs Peru, February 2, 1981 "**History *will never forget* this shameful betrayal, because except for it Siad Barre would not have invaded us, he would not have mercilessly butchered defenseless workers, toiling masses, children, and old people.**" Mengistu Haile Mariam, Ethiopia vs Somalia, August 20, 1977
Threats	"If this dispute is not solved through the diplomatic efforts, . . . **we will not be able to tolerate, under any circumstances, the positioning of those missiles on Lebanese territory and we will do what we have to do.**" Menachem Begin, Israel vs Syria, May 7, 1981 "It is for South Africa to choose whether to live in peace or in war. **But we are not prepared for a cold war. We prefer an open war.**" Samora Machel, Mozambique vs South Africa, February 14, 1981

Table 3.1: Example statements by each category
Parts of the text from the relevant category are boldfaced. Those from other categories are italicized. The statements in the "Threats" category describe only threats with no anger expressions.

relations faces significant external validity problems regarding its findings, the benefits of focusing on public statements outweighs the costs of excluding private statements.

3.2 Data Collection Procedure

I collected public statements directed at adversaries in the context of bilateral international crises identified by the International Crises Behavior data (Brecher and Wilkenfeld 1997). I focus on bilateral crises because my theory concerns dyadic interactions in which a challenger sends a threat a defender. The unit of analysis is a directed crisis dyad, where State A is the challenger who initiated the crisis and State B is the defender. To construct the dataset, I first identified interstate dyads that engaged in hostile actions against each other using the ICB's Dyadic-Level Crisis Data v. 14. I included intrastate crises that escalated into the international level as they involve international crisis participants. I then identified the challenger and the defender of each crisis using the *trigent* variable in the ICB's System-level Data v. 14, which indicates the entity that has triggered the crisis. When the crisis is triggered by a nonstate actor or an internal actor, or if the identity of the challenger is not clear, I examined the crisis start dates in ICB's corresponding actor-level dataset and coded the first mover as the challenger. For cases in which both actors have the same start dates, I consulted the ICB's case descriptions to identify the first mover and manually coded the challenger.[2] Finally, I merged this data with the Archigos leader data to record information about the leader(s) who held power during the course of a crisis (Goemans, Gleditsch and Chiozza 2009). This produced a dataset of 184 dyadic crises between years 1946-1996 involving 213 leaders.[3]

After identifying the relevant crises and leaders, I collected leaders' public statements. The public statements data are drawn from two different sources. First, I use the Foreign Broadcast Information Service (FBIS) to collect statements made by non-U.S. leaders. The FBIS was an intelligence program run by the Central

[2]There are six such cases.

[3]The dyadic data and variables are generated using *peacesciencer* R package by Miller (2021) (Version 1.0.0).

Intelligence Agency, which transcribed public statements issued through radio and televised broadcasts and press agency transmissions from almost every country for years between 1945-1996. The intelligence community used this information to "gauge local reaction to events ... [or] to support estimates of future events or to identify rhetorical patterns or broadcast schedules to support intelligence analysis" (Leetaru 2010). FBIS has two main advantages in that its spatial coverage is nearly global, and that its documents are translated into English, facilitating cross-national analysis without language barriers. Previous work also took advantage of these features of FBIS records (Shelef 2016; Katagiri and Min 2019). Because FBIS did not cover the U.S., I obtained statements from U.S. presidents from the Public Papers of the Presidents of the United States, which is a comprehensive database of all presidential statements collected by the Office of the Press Secretary since 1929 (Woolley and Peters 1999). The spatial scope of the final statements data is global, and the temporal scope is limited to the years 1946-1996 because of the limited availability of the FBIS records.

I collected statements at the paragraph level because a paragraph likely represents a topic, making it a natural unit of observation. While FBIS and the Public Papers provide access to various types of statements, I focused on spoken statements personally delivered by the leaders involved in the crises. This means that I excluded summary reports of speeches, news coverage of statements, written statements, and statements read by other political figures. I made this decision to avoid conflating increased media attention with a higher level of expressed anger. I identified crisis-relevant leaders using the Archigos data. To only record leaders who had a meaningful impact on a crisis, the dataset included a leader only if he or she was in power at least for a week over the crisis' duration. In cases where there was a leadership change during a crisis, I collapsed all statements made by all leaders who held power. This decision is theoretically motivated. Statements

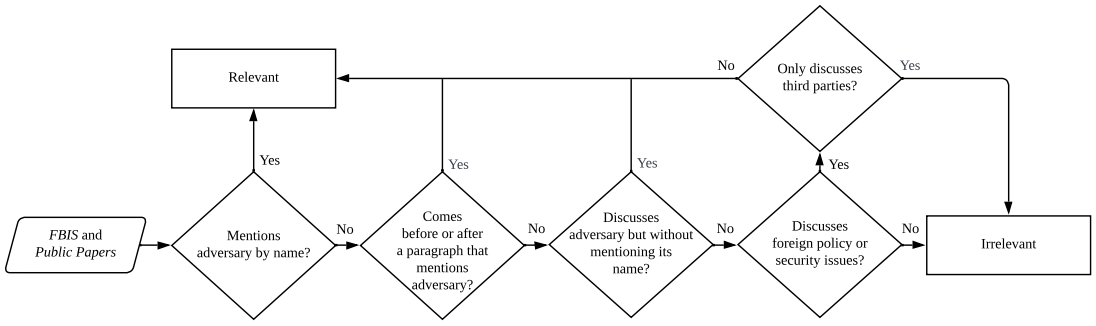

Figure 3.1: Decision process for collecting statements

from a previous leader could have lasting effects on an adversary's perceptions about the country's resolve.

Figure 3.1 illustrates the decision process for collecting relevant statements. I first searched the FBIS and Public Papers databases using a crisis participant's country name and leader name to download all available statements from crisis leaders. Online FBIS records provide images of the original transcripts, which are processed using optical character recognition (OCR).[4] I used the search function on the Public Papers database. I then retained only "relevant statements," defined as public statements delivered during the crisis or 30 days prior and addressing the adversary in the context of the crisis.[5] Relevant statements are defined as a paragraph that is made between the start and end dates of a given crisis at the system level and explicitly mentions the adversary using its country name, leader, or capital city, and paragraphs that come immediately before and after that paragraph.[6] I then continued reading subsequent paragraphs to determine

[4]However, a significant number of images were of such poor quality that extracting accurate text from them was not possible. I read and typeset all the downloaded statements to ensure accuracy.

[5]This approach is similar to that of McManus (2017).

[6]The ICB data record separate start and end dates for each crisis participant because the trigger and the termination of a crisis are measured based on the leader's perceptions. The system-level start dates record the earliest date a crisis is perceived by a leader, and end dates record the date when the last actor perceived that the crisis terminated. The use of the system-level variable is justified because leaders are likely to pay attention to their adversaries' behavior even after the crisis is initially deemed over.

whether they are also about the adversary. This step was necessary because leaders sometimes discuss adversaries and their actions using pronouns or without mentioning their names (e.g., using passive voice). Next, of those paragraphs that do not explicitly mention the adversary, I dropped the ones that are completely irrelevant to foreign affairs, such as greetings, procedural words, and remarks about domestic political candidates. In cases where a paragraph was about foreign affairs but only discussed a country that is not a party to the crisis, I assessed the relevance of the paragraph based on whether the other country was explicitly mentioned alongside the adversary in the preceding or following paragraph. This procedure resulted in a corpus of 5,807 public statements.

Finally, it is important to be transparent about the limitations of the data. One limitation is that I focus on the dominant leaders defined by Goemans, Gleditsch and Chiozza (2009). This choice admittedly led to the exclusion of statements from other political elites, such as the Chinese Premier Zhou Enlai's 1950 warning against U.S. incursions into North Korea. I focus on dominant leaders because including lower-ranking officials risks arbitrarily determining who is relevant, let alone the logistical challenge of collecting statements from all elites across multiple states. For instance, foreign and defense ministers might be obvious candidates, but there is no systematic guidance for including or excluding certain officials beyond intuition. In addition, although elite statements may generally influence crisis bargaining, statements by the heads of states should carry the most weight as they have the greatest decision-making power in foreign affairs, especially regarding the use of force. Another limitation of my dataset is that it only includes statements that were radio or television broadcast. Theoretically, however, this does not pose a significant concern since public statements that are only delivered locally would have a low chance of reaching an adversary.

Relatedly, the fact that my statements data spans 1946-1996 might raise con-

cerns about generalizing any findings from it to today's international crises, as recent technological innovation might have changed the effectiveness of elite emotion communications. While this is a valid concern, I contend that the media environment is more conducive for elite emotion communication in the present than in the past. Not only do politicians have many more opportunities to express emotions due to increased traditional media coverage, but new digital media might also better capture and disseminate politicians' spontaneous, fleeting emotional expressions in their statements, thanks to advances in information and communications technologies. Additionally, leaders may issue statements through informal medium such as social media. Recent work by Harris and Lin-Greenberg (2024) suggests that statements issued through social media are not seen as "cheaper talk" than statements delivered via a more traditional medium. In sum, there is both theoretical and empirical reason to expect that findings from my statements data will generalize to today.

3.3 Labeling Anger Expressions and Threats

Having described how I collected crises-relevant statements, I now turn to the coding procedure. While previous work that used text data employed automated approaches such as dictionary and supervised machine learning (McManus 2014; Schub 2022; Katagiri and Min 2019), I hand-labeled the each statement for anger expressions and threats. I made this decision because, despite recent advances in machine learning methods, careful human coding may remain the gold standard (Grimmer and Stewart 2013). With a small enough sample size of 5,807 statements, hand-coding is feasible and perhaps even desirable since training accurate machine learning models typically requires a larger sample size.

3.3.1 Measuring anger

Prior to reading the statements, I drew on the psychological literature on emotions to develop a theoretically grounded coding scheme for anger expressions. The first step involved making assumptions about how humans verbally represent discrete emotions. I adopted a well-known model of eight basic emotions consisting of anger, disgust, sadness, surprise, fear, trust, joy, and anticipation (Plutchik 2001). This model posits that each basic emotion has varying levels of intensity, and complex emotions can be represented as a combination of two basic emotions. For example, annoyance is low-intensity anger, and rage is high-intensity anger. Anger and disgust jointly produces contempt; anger and anticipation combine to generate aggressiveness. For my purposes, I only consider anger as a basic emotion and do not automatically label complex emotions that involve anger as anger itself.

I drew on the psychology literature to create four categories of verbal expressions of anger: *direct expressions*, *indirect expressions*, *verbal assaults*, and *accusations of wrongdoing*. These are motivated by theories about how humans express anger (Averill 1982; Sell et al. 2017) as well as existing measure of "anger-out," the tendency to express anger overtly (Deffenbacher et al. 1996). First, direct expressions of anger involve an explicit use of anger nouns, adjectives, and adverbs (e.g., "indignation," "outraged," "bitter," "angrily"). Anger can also be conveyed more indirectly through phrases and metaphors that require an contextual understanding to infer the implied emotion (Koschut et al. 2017). Consider, for example, the following passage from the statement by Indonesia's president Sukarno during a crisis with Malaysia. Parts of the text that indicate anger are italicized.

> "I now feel that *I have reached the limit of my patience*, and now I turn over *the crushing of Malaysia* to the Indonesian people. Yes, brothers, *I have reached the limit of my patience*, and now I turn the issue over to the

Indonesian people."[7]

This passage contains an indirect expression of anger that requires an understanding of the whole phrase. In addition, the word "crushing" conveys anger, compared to a choice of a more neutral word like "attacking." Accordingly, this statement indicates anger and an implicit threat to use force.

A *verbal assault* is another important category (Deffenbacher et al. 1996). It includes a broad range of rhetoric that includes "insults or the extreme use of profanity" (Spielberger et al. 1999), and general "negative characterizations" of the adversary (McManus 2017). The last category is an *accusation* of wrongdoing directed towards an adversary, which is the most prevalent forms of anger expressions (Frank 1988; Hall 2015). For a paragraph to qualify as an accusation, it must indicate that (a) a instigator (b) committed a transgression voluntarily, and (c) the transgression was unwarranted. This variable captures discussions of various anger-eliciting events, such as betrayal, deception, and human rights violations, which are attributed to the adversary. To illustrate, consider the following statement about South Africa from Angola's president Agostinho Neto:

> "The imperialists carry out diplomatic maneuvers and draw up many plans with a view to achieving false and peaceful settlement while they arm and approve *racist attacks [by the South African regime] against the peoples of southern Africa.* We now know their ways: After each *barbarous act of aggression by the racists*, the Westerners manifest their concern by launching a new diplomatic offensive and spreading falsehoods together with their *racist proteges*. Meanwhile, time is wasted and *the peoples of southern Africa bury their dead with suffering and humiliation*. This was what happened at Cassinga."[8]

[7] Sukarno, "Sukarno Issues Action Command at Bally," 03 May 1964.

[8] Agostinho Neto, "Neto Addresses Independence Anniversary Rally in Luanda," 11 November 1978.

This example has multiple components of anger. The statement accuses South Africa of the killing of civilians at the town of Cassinga, Angola, and verbally assaults South Africa by calling it ”racist” and ”barbarous.” Also note that I do not automatically label statements that include insults and anger words as expressing anger. I meticulously follow the context of the statement—defined as the paragraph being annotated and the paragraphs before and after that paragraph—to understand to whom the expressed anger is attributed. In this example, the ”imperialists” who are being accused of making ”false and peaceful settlements” refer to the Western powers. In addition, I do not code a statement as expressing anger if the anger does not have a target or if the identity of the target cannot be ascertained.

These four categories may not be exhaustive but cover a wide range of verbal expressions that research has shown humans employ to communicate anger. As the above examples illustrate, a statement can include more than one category of anger expression. Furthermore, I do not code threats and coercive demands as expressions of anger. While a threat of force would indicate anger in social interactions, in international politics, threats should be regarded a declaration of policy intent.

3.3.2 Measuring threats

I adopt a relatively broad definition of threats: statements spoken by state leaders to influence an adversary’s beliefs about their resolve, or willingness to use force. By threats, therefore, I am referring to statements that make specific promises to use military force as well as statements that only indicate demands, refusals, and policy stances without such promises. This breadth is justified because a statement can be a signal as long as it establishes a commitment to a position and a possibility of entering a war over the issue at hand. My definition is very similar

to the concept of "statements of resolve" (McManus 2017), which are "public statements which indicate that a country is committed to a position."

I coded whether each statement indicates willingness to fight or continue fighting. If a statement only uses ambiguous language hinting at using force without specifying a target, I coded it as indicating a threat only if the preceding and the following paragraphs both include a non-ambiguous threat. I did not label a statement as a threat if it simply describes a military action the speaker took in the past without linking it to future intentions.[9] To illustrate how I measured anger independently of threat, consider the following two passages. The parts of the text indicating anger are italicized, and the parts of the text conveying threats are bold-faced.

> "We have waited until they realize *what an injustice they are doing to us*. Now *we look indignantly at this attempt*. We are a peaceful country. We know what a clash would mean. Therefore we said that we renounce Triste because of peace and because we want cooperation."[10]

> "Armenia founded a national army a long time ago. This army has well-trained forces and is equipped with modern technology. **If we are unable to resolve the Nagomo-Karabakh conflict politically, we have no alternative but to protect our citizens and our territory.**"[11]

The first statement is from Yugoslavia's President Josip Tito during a crisis surrounding Triste. While Tito expresses anger at Italy about "an injustice they are doing," he suggests his willingness to make a concession to avoid conflict. In the second example, coming from Yagub Mammadov, an acting President of

[9] For example, saying "Our forces repelled enemy attackers" is not a threat, but adding "and we will do it again" at the end of that sentence turns it into one.

[10] Josip Tito, "Tito Interviewed Prior to Speech," 28 November 1953.

[11] Mamedov, "Mamedov on Solutions to Nagorno-Karabakh 'War,'" 3 March 1992.

	A1 Anger	A1 Threat	A2 Anger	A2 Threat
Accuracy	0.90	0.94	0.83	0.90
F1	0.91	0.94	0.90	0.94

Table 3.2: Inter-coder reliability between the author and Annotator 1 (A1) and Annotator 2 (A2)

Azerbaijan during a crisis with Armenia, illustrates an unemotional threat to use force.

I labeled the statements with two undergraduate student annotators, who worked independently after an hour-long training session. Validating my coding is crucial not only because of the concern that I was aware of my own hypotheses when labeling the data, but also because there might be room for subjective judgments when defining and detecting emotions in texts. To address these two issues, while the annotators were instructed to carefully follow the coding scheme discussed in the main text, they were also allowed to consider their intuition about a given text. Differences in labels were resolved through weekly discussions, but ultimately, the annotators decided whether to revise their own coding based on the discussion.[12] We produced three sets of labels and achieved high inter-coder reliability. I report accuracy and $F1$ scores in Table 3.2.

The labeling was done in a binary fashion. The variable *threat* takes a value of 1 if the paragraph indicates a willingness to use military force against the adversary and 0 otherwise. The variable *anger* takes a value of 1 if at least one of its sub-variables—*direct expression*, *indirect expression*, *verbal assault*, and *accusation of wrongdoing*—is 1, and 0 otherwise. In summary, the sub-variables are defined as follows:

- *Direct expression*: 1 if an emotion word is used to describe the speaker's anger regarding the adversary or the adversary's behavior; 0 otherwise.

[12]The discussions were one-on-one, and the two annotators did not work together.

- *Indirect expression*: 1 if anger is indicated through metaphors and/or phrases but without the use of emotion words; 0 otherwise.

- *Verbal assault*: 1 if the speaker verbally attacks or degrades the adversary or the adversary's behavior; 0 otherwise.

- *Accusation of wrongdoing*: 1 if the speaker accuses the adversary of a wrongdoing; 0 otherwise. The paragraph needs to have information that (a) the instigator (b) voluntarily committed an (c) unwarranted transgression. This includes betrayal, violating agreements, lying, human rights violations by the adversary. The victim of the transgression does not have to be the speaker's country.

3.4 Data Description

Using the statements data, I created scores for anger expressions and threats for the challenger and the defender in a given dyadic crisis. The scores are defined as the proportions of statements that indicate anger or threats out of the total number of statements the leader has issued over the course of the crisis. The measures are therefore bounded between 0 and 1, where 1 indicates that every statement includes an expression of anger or a threat, and 0 means that the leader made at least one public statement about the adversary but neither communicated anger nor issued a threat. The scores for leaders who did not make any public statement are coded as zero rather than missing because the decision not to issue any public statement is analogous to the decision not to publicly express anger or threaten force. In Chapter 4, I include a binary variable indicating whether the crisis leader made any public statement in the statistical analysis to account for the independent effect of making public statements.

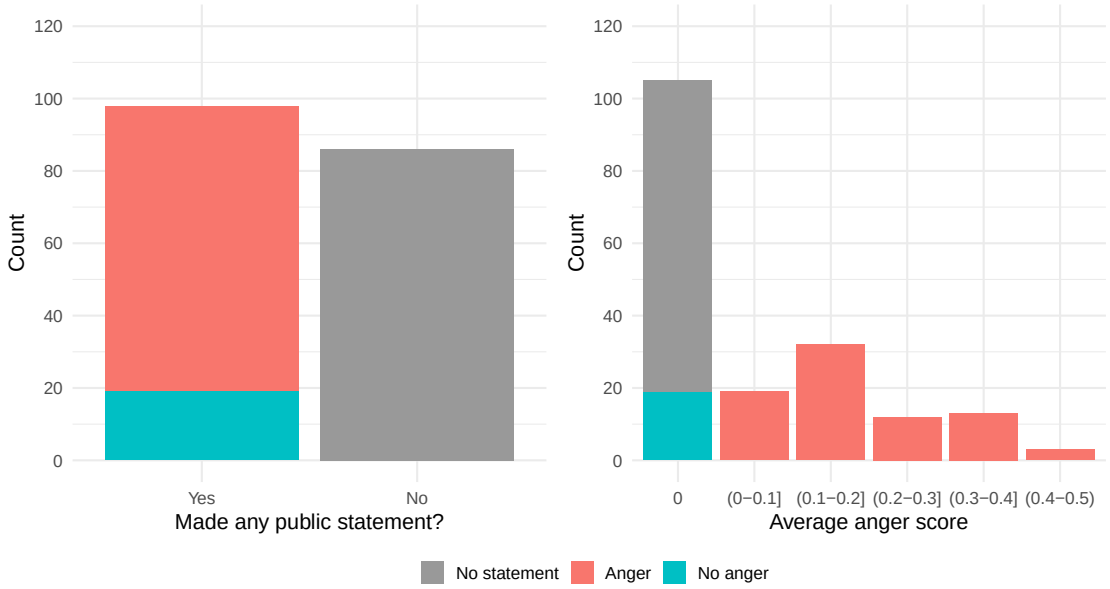

Figure 3.2: Distributions of anger scores aggregated to the crisis level
The no-statement cases are color-coded as gray. The red bar represents crises where at least one statement indicates anger, and the blue represents cases where no statement indicates anger. The left panel describes how many crises involve at least one public statement from either crisis participant, and how many of those crises involve any angry statements. The right panel illustrates the distribution of anger scores using the same data. Each bin represents a 0.1 interval except for the "0" bin. For presentation purposes, the scores are averaged between the challenger and the defender, treating no-statement cases as 0. The figure suggests that there is considerable variation in public statements and anger expressions during crises.

Figure 3.2 plots the distribution of anger scores at the crisis level. The gray bar represents crises in which no public statement is issued. The red bar represents crises where at least one statement indicates anger, and the blue represents cases where no statement indicates anger. The left panel describes how many crises involve at least one public statement from either crisis participant, and how many of those crises involve any angry statements. The right panel illustrates the distribution of anger scores using the same data. Each bin represents a 0.1 interval except for the "0" bin. For presentation purposes, the scores are averaged between the challenger and the defender, treating no-statement cases as 0.

The left panel suggests that public statements are not common, but most of

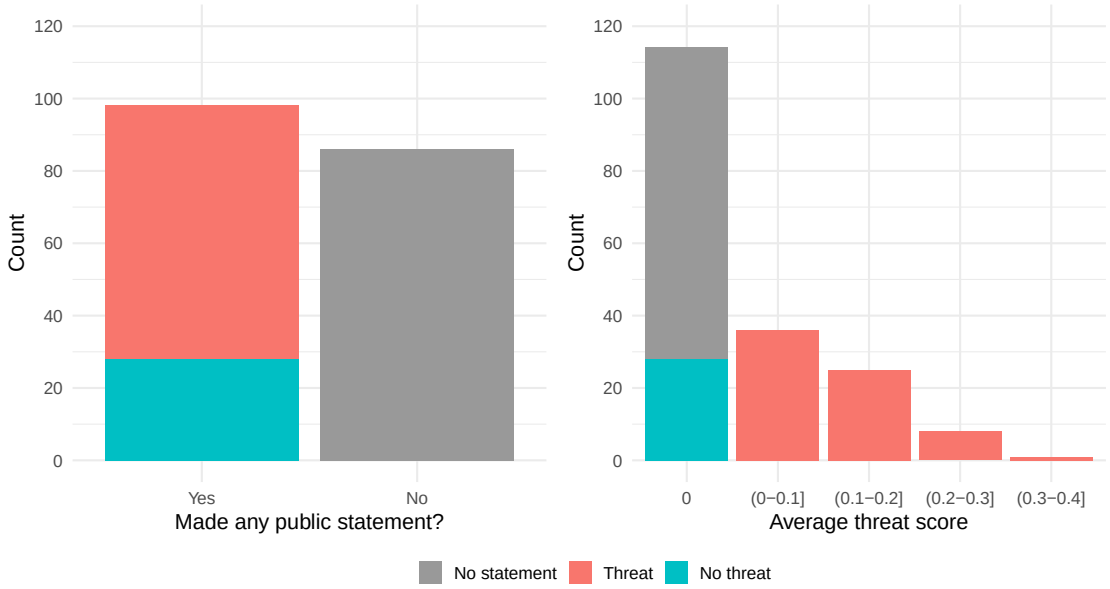

Figure 3.3: Distributions of threat scores aggregated to the crisis level
The no-statement cases are color-coded as gray. The red bar represents crises where at least one statement indicates anger, and the blue represents cases where no statement indicates anger.

the leaders who make public statements express anger. About 53% of the 184 bilateral crises involve at least one leader broadcasting a public statement; in about 81% of those crises, the leader expressed anger toward the adversary. The right panel provides a more disaggregated view of the distribution by plotting the average anger scores between the challenger and the defender in each crisis. This plot suggests that there is significance variation in the intensity of anger expressions. The figure also shows that the distribution is skewed, with about 57% of the observations having a score equal to zero. However, the skewness is driven by cases with no statements rather than actual instances of no-anger statements. This is in keeping with the intuition that public statements should be relatively infrequent as they are costly (Fearon 1997).

Figure 3.3 displays the distribution of threat scores in the same way as Figure 3.2. Similar to the anger expressions, the left panel suggests that most public statements include threats, although it appears that statements contain fewer

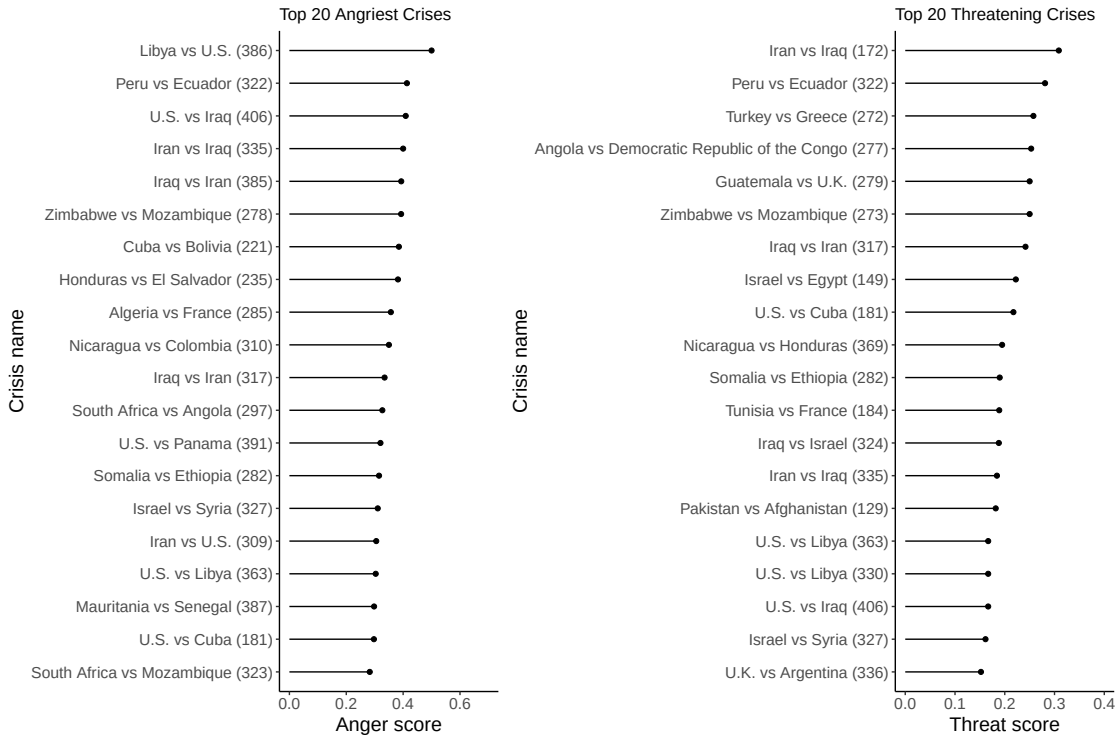

Figure 3.4: Twenty crises with the highest anger scores and threat scores

threats than anger expressions. The right panel shows significant variation in threat scores, with the similar skewness as anger scores due to no-statement cases. Additionally, the threat scores appear to be clustered at the lower end of the distribution with the mean of 0.085, suggesting that on average, leaders issue threats moderately. Indeed, if public threats are costly signals that are strategically employed (Fearon 1994; Sartori 2005) and anger is an index that is "cheap" but difficult to manipulate (Jervis 1970), it would be unsurprising to observe more anger expressions than threats.

Next, Figure 3.4 plots the twenty crises with the highest anger scores (left panel) and threat scores (right panel). Both panels show that the scores vary considerably in size without apparent outliers. It is also worth noting that there is a 40 percent overlap between the two panels. Qualitatively examining the top cases in each panel reveals that they involve significant violence and harsh accusations and

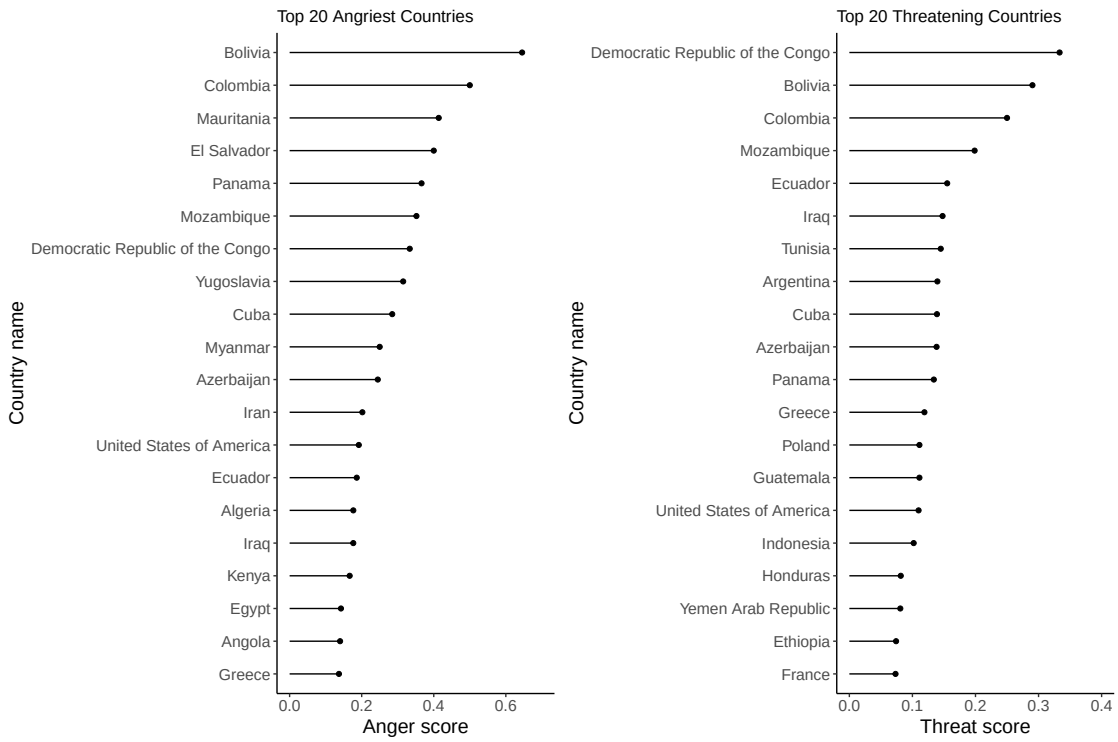

Figure 3.5: Twenty states with the highest anger scores and threat scores

verbal assaults against the adversary. For example, the crisis between the U.S. and Libya in 1988 (crisis no. 386) was triggered by Libya's alleged crashing of a U.S. commercial plane that caused the deaths of over 250 civilians, as well as by U.S. intelligence that Libya had begun the production of chemical weapons. The U.S. later shot down two Libyan jets over international waters. The second highest case (crisis no. 322) begun with an armed clash between Peru and Ecuador in 1981 that quickly escalated into border closures, mobilization, and suspension of commercial flights. On the right panel, the crisis with the highest threat score (crisis no. 172) involves multiple border disputes and skirmishes over a joint waterway between Iran and Iraq. Overall, crises with high anger and threat scores fit the historical narrative of violent militarized crises, lending face validity to the coding of statements.

Finally, Figure 3.5 displays the twenty states with the highest anger scores (left

panel) and threat scores (right panel). We see that states issue varying levels of anger statements and threats, and there are no high outliers. While some states appear to rank higher than other states in both the anger and threat scores, they are involved in fewer bilateral crises than other states. For example, Bolivia has the highest anger score and the second-highest threat score but appear only once in my data, in a crisis against Cuba triggered by Cuban revolutionary Ché Guevara's guerilla activity in 1967. Similarly, while Colombia ranks second in the anger score and third in the threat score, it engaged in a single crisis with Nicaragua in 1979 after Nicaragua declared jurisdiction over what Colombia perceived as its territory. The two crises that Bolivia and Colombia were involved in appear in Figure 3.4. Compare these states with the U.S., which initiated seven crises and became a target in ten crises but ranks outside of the top ten in the anger or threat scores. Therefore, we may interpret Figure 3.5 as describing participants of particularly anger-inducing or intense crises, and inferences about particular leaders' tendency to threaten or express anger should be made with caution.[13]

[13]Additionally, it should be noted that my data only comprise of bilateral crises. So this analysis cannot offer a complete picture of anger expressions from individual states.

Chapter 4

Cross-National Analysis of Public Threats and Anger Statements

Evaluating the hypothesis that threats lead to higher chances of winning in international crises when they are accompanied by greater levels of anger expressions requires two types of data. First, we need variation in different types of international crises across time and space. Second, we need variation in threats and anger expressions in statements leaders made with respect to these crises. While there are existing datasets on international crises and disputes (Brecher and Wilkenfeld 1997; Palmer et al. 2022), as well as datasets on U.S. presidents' public statements (McManus 2017), we currently lack on comparable data on leaders' emotional expressions on a global. To address this shortcoming, I built a new dataset on public statements by hand-labeling anger and threats in public speeches from world leaders involved in international crises between 1946-1996. The previous chapter provided some descriptions of the data and explored the conditions under which leaders are likely to make anger statements against their crisis adversaries. This chapter uses the data to conduct statistical analyses to establish correlational evidence for anger expressions.

4.1 Research Design

4.1.1 Dependent Variable

This chapter investigates whether anger improves crisis outcomes by making threats more credible. Naturally, the dependent variable in the analysis is the crisis outcome. The crisis data are drawn from the International Crisis Behavior (ICB) project (Brecher and Wilkenfeld 1997). One advantage of the ICB dataset is that it includes a broad spectrum of crisis behavior, such as economic sanctions and arms transfers to the adversary. In contrast, other datasets, such as Militarized Interstate Disputes (MIDs) (Palmer et al. 2022), focus on disputes involving an explicit threat or use of force. To test my theory across a wider array of crisis situations, I use the ICB data set. While anger may be particularly relevant to militarized crises, which might implicate great human and economic costs, the emotion of anger has evolved to regulate various types of social interactions, not just violent ones. I thus expect that anger expressions will influence the credibility of threats in non-militarized crises.

The dependent variable in this study measures the degree to which the challenger is able to secure the interests at stake in the crisis. I created the variable drawing from the ICB, which codes crisis outcome based on four categories: victory, compromise, stalemate, and defeat. In keeping with established practices in the literature (Gelpi and Griesdorf 2001; McManus 2017), I constructed a three-point scale by merging the compromise and stalemate categories. The dependent variable is coded as 3 if the crisis outcome is a clear victory for the challenger, 2 if the outcome is a compromise or stalemate, and 1 if the outcome is a clear defeat for the challenger.

4.1.2 Independent Variables

As described in detail in the previous chapter, the main independent variables for the analysis are the threat and anger scores of challengers and defenders in bilateral crises. These scores are measured as the proportion of threatening or angry statements out of the total number of statements the leader has issued in the crisis. Since my hypothesis posits anger as a mechanism for threat credibility, the natural quantity of interest in the statistical model is the interaction of the two variables.

The analysis's identification strategy relies on selection on observables. While it would be extremely challenging to eliminate omitted variable bias, I include a set of covariates both at the state level and the crisis level to rule out the most obvious alternative explanations. First, one might be concerned that due to subjectivity involved in interpreting emotions, my coding of anger expressions might inappropriately pool different categories of emotions, capturing strong emotionality rather than the specific anger emotion. Thus, I include measures of emotionality developed using the NRC Word-Emotion Association Lexicon (NRC) (Mohammad and Turney 2013). The NRC lexicon contains over 10,000 words labeled with eight "basic" emotions—anger, sadness, joy, fear, disgust, surprise, trust, and anticipation (Plutchik 2001). I applied the dictionary to the corpus to count the instances of each emotion category in each statement. The final measure of emotionality is defined as the proportion of emotional words to all words spoken by a leader in a crisis.

If going public with threats signals resolve by tying the leader's hands (Fearon 1997), any public statement regardless of its emotional content might affect crisis outcomes. To address this concern, I include a variable indicating whether or not the leader issued a public statement in the crisis. Moreover, anger expressions might simply indicate issue salience or high interests. It could be that leaders

express greater anger with regards to crises they believe are more important, and simultaneously bargain harder to achieve more favorable outcomes. Following McManus (2017), I include a daily word count as a measure of the salience of a crisis, assuming that leaders would speak at greater lengths about events they care more about. Additionally, I account for whether a state experienced a leadership change during a crisis because a new leader might introduce new speech patterns and increase uncertainty about resolve (Smith and Spaniel 2019).

Audience cost scholars suggest that leaders who can be politically vulnerable are better able to signal resolve through public statements and win crises, compared with leaders who are insulated from political pressure (Schultz 2001). If this is the case, any observed relationship between anger expressions and favorable crisis outcomes might be plausibly attributed to potential political vulnerability. To rule out this explanation, I use regime type data from Geddes, Wright and Frantz (2014) and include a dummy variable indicating whether the state is a personalist regime. Personalist regimes are the only regime type believed to be unable to generate audience costs (Weeks 2008).[1]

Next, I address potential crisis-level confounding. Research suggests that in social interactions, powerful individuals are more likely to experience and express anger than powerless individuals (Sell, Tooby and Cosmides 2009; Petkanopoulou et al. 2019). If this dynamic holds among political leaders, those with military superiority might be more inclined to express anger at their less powerful adversaries and have a higher likelihood of winning. I therefore include a measure of relative power using the National Material Capabilities data (Singer 1988), defined as the Challenger's relative share of total capabilities in a dyad. Furthermore, higher

[1]Substituting this variable with a measure of democracy, defined as having a Polity2 score higher than 6 (Marshall and Jaggers 2002), does not change the results. I used the regime type information of a state for the year when the crisis is terminated. The GWF data exclude countries with small population. For Guyana, Solomon Islands, and Papua New Guinea, I used the Autocracies of the World dataset (Magaloni, Chu and Min 2013).

levels of violence could intensify a leader's anger expressions and simultaneously shape the crisis outcome. I account for the highest level of violence employed during a crisis, which could be no violence, minor clashes, serious clashes, and a full-blown conflict (Brecher and Wilkenfeld 1997). I also address a concern that certain issues may elicit greater anger from leaders and simultaneously motivate them to bargain harder. It is plausible that security issues implicate national interests to a greater extent than non-security issues. I include a variable in the model indicating whether the crisis involves security or military matters or not. Lastly, I account for crisis duration, which could affect the number of statements that can be made and other unobserved crisis processes.

4.2 Results

4.2.1 Anger Expressions Improve Threat Effectiveness

Testing whether anger expressions increase threats' effectiveness requires modeling crisis outcomes as a function of the interaction between the two variables. Given the ordinal nature of the dependent variable, I use ordered logit regression. In all models, robust standard errors are clustered at the dyad level because the same dyad can engage in recurring crises. The results reported in Table 4.1 support Hypothesis 1. Model 1 describes the parsimonious model only including interactions between the threat and anger scores, along with their constitutive terms. Model 2 includes the full set of covariates. The interaction coefficient for State A is positive and statistically significant in both models, suggesting that angrier threats are more effective at persuading the defender to make concessions. On the other hand, although the same interaction coefficient for State B is negative, it did not achieve statistical significance at the conventional level. This pattern is consistent with the scope condition that anger will be more useful for challengers

	Model 1	Model 2
Anger, A × Threat, A	36.525 (15.245)**	35.675 (12.065)***
Anger, A	0.606 (1.621)	4.606 (2.297)**
Threat, A	−3.513 (2.967)	−0.903 (2.920)
Anger, B × Threat, B	−11.142 (9.032)	−3.191 (13.646)
Anger, B	−1.469 (1.262)	−3.286 (2.000)
Threat, B	0.246 (2.773)	−5.007 (4.886)
Statement, A		1.499 (1.308)
Statement, B		1.762 (1.182)
Emotionality, A		−10.981 (7.211)
Emotionality, B		−9.503 (6.694)
Word count, A		−0.009 (0.004)**
Word count, B		0.004 (0.002)*
Personalist, A		−0.784 (0.498)
Personalist, B		0.488 (0.480)
Leadership change, A		0.858 (0.470)*
Leadership change, B		0.642 (0.540)
Relative power, A		0.824 (0.582)
Level of violence		0.110 (0.176)
Crisis duration		−0.001 (0.001)
Military issue		0.321 (0.451)
1—2	−1.073 (0.240)***	−0.394 (0.596)
2—3	1.671 (0.237)***	2.740 (0.678)***
Num. obs.	184	179

***$p < 0.01$; **$p < 0.05$; *$p < 0.1$. Robust clustered standard errors at the dyad level.

Table 4.1: Interaction effects of anger expressions and threats on crisis outcomes The table shows the estimates of two ordered logit regression models. The full model has five fewer observations than the parsimonious model due to missingness in the regime type variable.

than defenders.

It is difficult to interpret regression coefficients of logistic models as they are reported as log-odds. To illustrate the substantive interaction effects, Figure 4.1 uses the estimates of Model 2 to plot the marginal effects of State A's threats on the probability of achieving winning outcomes relative to the other two outcomes, over the range of its anger scores.[2] Other covariates are held constant at their mean values. The figure describes how the credibility of threats changes at varying

[2]The analysis focuses on State A since State B's interaction coefficient is not statistically significant.

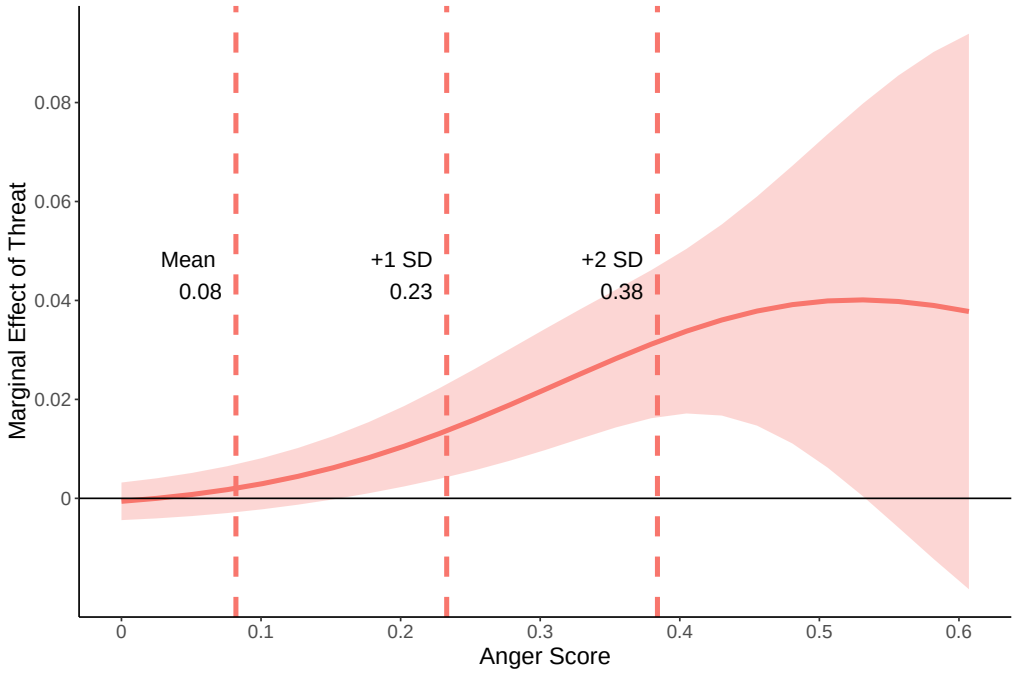

Figure 4.1: Marginal effect of threat on the probability of winning across anger scores for State A
The results for State B are not displays as they are not statistically significant. The dotted red lines indicate the mean anger score, and one- and two-standard deviation increases in the anger score.

levels of anger expressions. The upward slope of the line confirms the hypothesis that the effect of threats on the probability of winning increases as the threatening leader's anger increases. Examining the plot, a one-standard deviation increase from the mean anger score appears to be associated with about a 1 percentage point increase in a threat's ability to achieve a winning outcome. The magnitude of the association is relatively small, but it still suggests a significant role of anger expressions, given that leaders can gain this benefit by verbally communicating anger rather than taking material actions.

In addition, while the confidence band stays above the zero line at high levels of the anger score, it overlaps with the zero line until the anger score reaches 0.15, roughly the midpoint between the mean and one standard deviation to its right. This suggests a striking result: public threats may fail to convey resolve unless

leaders also express a substantial amount of anger. The marginal effects again become indistinguishable from zero at extreme levels of anger, likely due to higher uncertainty resulting from the few data points in this part of the distribution. Overall, the results indicate strong evidence for the contention that anger is a key mechanism for threat credibility.

4.3 Addressing Confounding

The preceding section has provided evidence for the role of anger expressions. Yet, anger expressions do not arise at random. Like any other observational study of strategic behavior, the results of the current study are subject to systematic bias due to selection effects, reverse causation, and confounding by third factors. Furthermore, given that my statements data are drawn from the former CIA intelligence program, there could also be concerns about data collection bias within the Foreign Broadcast Information Service (FBIS) due to the agency's policy mandate. This section identifies possible sources of confounding and presents a series of supplementary analyses to address them both qualitatively and quantitatively, demonstrating the robustness of the findings. In order to show that the results are not driven by specific coding decisions, it also replicates the main statistical analysis with different model specifications and alternative measures of the independent variables.

Perhaps the most obvious candidate to investigate is a selection issue. If expressions of anger carry strategic benefits as my theory predicts, and if experiences of anger simultaneously increase willingness to fight (Hall 2017; Powers and Altman 2023), then leaders might express anger at greater rates while fighting or bargaining more aggressively. This implies a challenge for the analysis, in which we are unable to isolate the informational effect of anger from its behavioral

effect.[3] The positive relationship between anger expressions and the probability of winning might be attributed to the underlying resolve and actual bargaining behavior. This issue could be re-framed as a type of reverse causation. Instead of anger expressions improving threat credibility and thus the likelihood of winning, leaders who anticipate winning might lash out at the adversary to appeal to the domestic public.

I meet these challenges by rejecting hypotheses that we expect to be confirmed if the aforementioned issues are true. First, we would expect that predictors of winning outcomes, such as relative power, will be correlated with expressions of anger. I estimate a tobit model predicting the anger scores of each crisis dyad member, regressing relative power on anger scores. I use a tobit model because the anger score is bounded between 0 and 1. The model accounts for other factors that might influence the propensity to express anger, such as the adversary's anger, rivalry (Thompson and Dreyer 2011), regime type, level of violence, foreign policy interest similarity (Chiba, Johnson and Leeds 2015), the presence of third party mediation, and whether the crisis involves a military issue.

Table 4.2 shows that relative power is not statistically significantly correlated with anger expressions. This result provides evidence indicating that leaders who anticipate winning or intend to stay firm do not necessarily express greater anger than those who do not. On the other hand, other variables appear to be significantly correlated with anger scores in ways we might expect them to be in the real world. Anger expressions from the crisis adversary are a particularly strong predictor of the leader's anger expressions. While this result is not surprising, it is still theoretically interesting as it suggests that public statements could be modeled

[3]The next chapter uses a survey experiment to randomize respondents' assignment into an adversary leader's anger expressions to estimate the impact of anger expressions on beliefs about the adversary's resolve. Although the experiment may not speak directly to actual crisis outcomes, combined with the correlational evidence presented in this chapter, it communicates the strong plausibility of my theory.

	Model 1
(Intercept)	$-0.260\ (0.120)^{**}$
Relative power, A	$-0.012\ (0.095)$
Adversary's anger	$0.522\ (0.150)^{***}$
Rivalry	$0.118\ (0.083)$
Interest similarity	$-0.523\ (0.135)^{***}$
Level of violence	$0.056\ (0.034)^{*}$
Military issue	$0.125\ (0.075)^{*}$
Mediation	$0.100\ (0.067)$
Crisis duration	$0.000\ (0.000)^{***}$
Total	176
Left-censored	133
Uncensored	43
Right-censored	0

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$. Robust clustered standard errors at the dyad level.

Table 4.2: Tobit models predicting anger scores for State A and State B

as an escalating dialogue between two leaders. Leaders might lock themselves into a costly conflict by expressing increasingly intense anger against their opponents (e.g. Fearon 1997).

It is also worth noting that whereas rivalry is not significantly associated with anger expressions, foreign policy interest similarity is and has a negative coefficient. One possible reason for of this outcome is that anger may antagonize its targets. Leaders may not discriminate between rivals and non-rivals upon entering a crisis, but they might be reluctant to further estrange opponents that have similar foreign policy interests and could be reconciled with in the future. In addition, crises that revolve around military and security issues involve higher levels of anger expression, in keeping with my expectation that such crises might incite greater anger by putting greater national interests at stake. Lastly, we see that level of violence and crisis duration are positively associated with anger expressions. These may be unsurprising results, since violent and protracted crises are likely to induce greater anger.

	Winning outcome	All outcome
(Intercept)	0.071 (0.103)	−0.034 (0.048)
Beginning	0.211 (0.109)*	0.222 (0.059)***
Middle	0.143 (0.112)	0.189 (0.064)***
Total	235	731
Left-censored	85	306
Uncensored	142	406
Right-censored	8	19

*** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$. Robust clustered standard errors at the crisis level.

Table 4.3: Tobit models predicting mean anger scores by crisis phase
The reference category is End.

The second way to address the selection and reverse causation issues is to analyze the timing of anger expressions. Leaders may become more certain about their probability of winning as the crises they are involved in approach the end. If the reverse causation is true, then we would expect to observe higher levels of anger expressions from winning leaders near the end of the crisis than at the beginning. Analysis of the timing of anger expressions suggests that this may not be the case. I created a subset of data that includes at least one public statement and divided the duration of each crisis in the subset into three phases equal in length—the beginning, middle, and end—and regressed the mean anger scores on them. Similar to the previous analysis reported in Table 4.2, I use a tobit model since the anger scores are defined as proportions. The model estimates are shown in Table 4.3. The reference category is the end of the crisis.

Examining the winning leaders' anger expressions in the left column, there do not appear to be statistically significant differences in anger expressions based on the timing at the conventional level. Although winning leaders might express anger at greater rates in the beginning compared to the end, the coefficient is significant only at the 10% level. The right column displays the results pooling all three outcomes: winning, losing, and compromising leaders. We again observe

that leaders do not communicate more anger at the end of the crisis. In fact, the coefficients of the Beginning and Middle variables are positive and significant, suggesting that upon entering a crisis, leaders might begin by expressing intense anger and then tone down their anger expressions as the crisis progresses towards the end. In sum, analyses of the relationships between proxies of winning and anger expressions and the timing of anger statements fail to provide evidence for selection effects or reverse causation. These results increase confidence that my findings are not biased by these issues.

Next, I address concerns regarding possible omitted variable bias. The main statistical model includes a battery of covariates to minimize this concern. However, there could still be confounders that are not accounted for in the regression in ways that bias the results away from the null. One such concern is the importance of anger statements relative to the use of force. The reasoning here might be that expressions of anger, or public statements in general, will provide little informational value in crises that involve the use of force because using force is arguably the strongest signal of resolve (or one might call it a realization of resolve). Granted, rationalist models of crisis bargaining tend to assume away the impact of force when analyzing threat effectiveness. However, as McManus (2014) explains, "the calculations involved in making the decision to fight and making the decision to continue fighting are very similar." One unique feature of force is that battle outcomes can alter future resolve and capabilities. In fact, recent scholarship has characterized war as a continuous bargaining process, in which negotiations take place between bouts of battles to recalibrate the opponents' willingness to continue fighting (Slantchev 2003; Filson and Werner 2002). Militarized crises could be characterized in the same way. Unless force is used in a way that completely neutralizes the adversary, there will remain uncertainty about the adversary's resolve. Therefore, leaders' statements and expressions of

anger would have room to make an impact of beliefs about resolve after force is used.

This might raise another concern about how often force was the primary determinant of the crisis outcome (e.g., forcibly removing the adversary leader). In such crises, the effects of anger expressions picked up by the regression model would be largely spurious. I consulted the ICB documentation and identified that out of 184 crises in my sample, the adversary was defeated predominantly by force in only six cases.[4] I re-ran the main analysis with a sample excluding these six cases, and, as shown in Model 1 of Table 4.4, the results are largely identical.

Because this study relies on historical data on public statements spanning five decades, it is possible that the use of anger words or the norms surrounding emotional expressions have changed over time. It might also be that the way leaders build and project their public images has changed in accordance with changes in the media environment. Model 3 in Table 4.4 demonstrates that the results do not change after I control for a time trend using the cubic polynomial approach suggested by Carter and Signorino (2010).

Relatedly, there might be concerns about the extent to which the results will generalize to more modern contexts. In fact, I anticipate that the findings will hold greater relevance to contemporary crises. Advances in communication and information technology have greatly increased leaders' media exposure. Today's leaders have more opportunities to reveal their emotions, either through public appearances or by "getting caught" in the media. Moreover, compared to the CIA reports from which my data is drawn, today's video and audio recordings are of much higher quality, making it easier for observers to detect emotions. Leaders' statements may also disseminate quickly and widely, making them more likely to

[4]The cases are: the Hungarian Uprising, the Guerrilla activity by Cuban revolutionary Ché Guevara in 1967, the Nicaraguan Civil War in 1978, the invasion of Grenada, a clash between Angola and Zaire (currently the Democratic Republic of the Congo) in 19877, and the invasion of Panama.

reach relevant targets.

The way people express anger might also differ across cultures and languages. Three issues might be worth discussing here. First, can anger be communicated effectively cross-culturally? Most individuals, including political elites, follow some display rules set by social and cultural norms, and this socialization of emotion expressions is believed to begin as early as infancy (Matsumoto 1990; Malatesta and Haviland 1982). Moreover, there may be cultural variations in the intensity of expressed emotions. For example, individuals from countries with more individualistic cultures tend to be more expressive than those from more collectivist ones (Matsumoto, Yoo and Fontaine 2008). This implies that the social meanings commonly attributed to particular emotion expressions in one country might not have the same impact in other countries. However, it is important to note that these issues do not necessarily invalidate the general conclusion that anger expressions influence crisis outcomes. Rather, they suggest that the effectiveness of anger expressions might vary depending on how culturally or linguistically similar the crisis adversaries are. I include a dummy variable indicating whether the crisis took place between states in the same geo-political region using the variable drawn from the Variety of Democracy data (Coppedge et al. 2023). Model 3 in Table 4.4 show that the results remain similar.

Second, I code anger based on statements that are translated to English from their original languages. This might raise concerns about the quality of the translation. For example, do words infused with anger in English convey similar emotions in other languages? It would be difficult to directly speak to the accuracy of the translation without cross-validating each of the transcript against its original document. Collecting and comparing thousands of historical documents in multiple languages will become intractable very quickly, and such a task is beyond the scope of this project. Instead, I will provide more background on FBIS

to illustrate its quality.

According to a declassified CIA document on the early history of the service, "[s]tandards of capability set for FBMS editors and analysts were very high,"[5] requiring that they not only hold a graduate degree in foreign affairs but also be "well informed, 'in a political sense,' on various countries" through first-hand foreign experience (Roop 1969). Similarly high standards were required for translators. At least in the beginning, FBIS considered "only those capable in at least two foreign languages ... with the additional requirement that they have some experience in foreign affairs or had resided in foreign countries." The translators were also highly educated, with about ten percent of them holding doctoral degrees. A contemporary assessment of FBIS's work suggests that its reports are "renowned for the extremely high quality of their translations, which often capture the tone and nuance of the original vernacular," which was achieved thanks in part to their "high level of editorial input, including iterative revision processes" (Leetaru 2010). In sum, the rigor of FBIS editorial efforts provides reassurance on the quality of the translation.

The third concern to address is that some degree of anger expressions might be missed while the statements were transcribed from audio or video formats to texts. The rigor of translation by the FBIS personnel, described in the preceding paragraph, reassures us that the transcribers paid particular attention to the tone of the statements. In fact, the main issue here is less about whether anger was lost in transcribing than whether I was able to pick up on anger from the transcriptions as intended by the transcribers. Indeed, a large body of research on facial expressions and emotion recognition suggests that detecting emotions may indeed be easier with visual or auditory inputs than with texts alone (Busso et al. 2004; Mehu et al.

[5]FBIS was called Foreign Broadcast Monitoring Service (FBMS) when it was formed in 1941 under the authority of the Federal Communications Commission. The service was renamed to FBIS after it was transferred to the CIA in 1946.

2012). Critically, however, this implies that I might under-measure anger rather than over-measure it, potentially creating a bias toward the null. The fact that my hypothesis has found statistically significant support therefore alleviates this concern.[6]

Finally, I address concerns about data collection bias, in which the FBIS data prioritize countries where the U.S. had strategic interests. Analyses of the geographic coverage of FBIS "do not appear to indicate that FBIS appreciably favored regions in which the United States was actively engaged during 1994-2004," even though Russia and China received greater attention than other countries (Leetaru 2010). While the report focuses on a different time period than my data, it provides a broad assessment that "FBIS provided a monitoring all regions of the world relatively evenly rather than a tactical resource focused on troublesome areas" (Leetaru 2010). It also notes that the FBIS coverage for African countries is lower than countries in other continents, but this is the result of underdeveloped media networks in African locations rather than lack of strategic interests.[7]

4.4 Using Alternative Measures of Anger Expressions

The previous section has addressed various confounding issues to demonstrate the robustness of the statistical results. In doing so, I paid particular attention to the quality of the translation and the text form of my statements data. In this section, I take a step further to address concerns about the coding of anger. As alluded earlier, I worked with two independent student annotators to tackle the concern that I was aware of my own hypotheses and crisis outcomes when I coded anger.

[6]It is theoretically possible that the null finding for the defender's anger and threat can be attributed to the under-measurement of anger. However, I doubt that this is the case, as there is little reason to believe that defenders' anger expressions were systematically harder to measure than challengers'.

[7]Theoretically, underdeveloped media networks could mean that it is hard for leaders to communicate public threats to an adversary.

	Model 1	Model 2	Model 3
Anger, A × Threat, A	40.753 (11.994)***	32.742 (12.602)**	37.343 (12.470)***
Anger, A	4.560 (2.322)*	5.358 (2.297)**	4.797 (2.363)**
Threat, A	−1.502 (3.064)	0.018 (2.964)	−1.020 (3.123)
Anger, B × Threat, B	−0.927 (14.372)	−6.459 (13.953)	−3.901 (14.322)
Anger, B	−3.302 (2.163)	−3.650 (2.003)*	−2.842 (2.103)
Threat, B	−5.273 (5.342)	−3.628 (4.929)	−5.606 (5.127)
Statement, A	1.383 (1.402)	1.453 (1.287)	1.539 (1.315)
Statement, B	1.954 (1.308)	2.061 (1.149)*	1.826 (1.256)
Emotionality, A	−10.161 (7.956)	−10.887 (6.975)	−11.638 (7.566)
Emotionality, B	−10.851 (7.508)	−10.702 (6.474)	−10.215 (7.138)
Word count, A	−0.010 (0.004)**	−0.011 (0.004)***	−0.009 (0.004)**
Word count, B	0.005 (0.003)*	0.005 (0.003)**	0.004 (0.003)
Personalist, A	−0.855 (0.488)*	−0.807 (0.510)	−0.885 (0.506)*
Personalist, B	0.413 (0.526)	0.607 (0.487)	0.469 (0.491)
Leadership change, A	0.953 (0.507)*	0.930 (0.481)*	0.892 (0.494)*
Leadership change, B	0.595 (0.555)	0.778 (0.527)	0.549 (0.543)
Relative power, A	0.711 (0.584)	0.833 (0.579)	0.971 (0.581)*
Level of violence	0.122 (0.181)	0.064 (0.173)	0.111 (0.175)
Crisis duration	−0.002 (0.001)	−0.002 (0.001)	−0.001 (0.001)
Military issue	0.195 (0.464)	0.134 (0.464)	0.189 (0.454)
Same region		0.918 (0.362)**	
t			29.477 (11.313)**
t^2			−122.796 (53.087)**
t^3			147.278 (70.022)**
1|2	−0.613 (0.587)	0.102 (0.645)	1.265 (0.787)
2|3	2.610 (0.678)***	3.304 (0.744)***	4.479 (0.841)***
Num. obs.	174	179	179

***$p < 0.01$; **$p < 0.05$; *$p < 0.1$. Robust clustered standard errors at the dyad level.

Table 4.4: Robustness checks using different model specifications
Model 1 uses the sample without cases where force was the primary determinant of crisis outcomes. Model 2 includes the region control. Model 3 accounts for time dependence.

Since the annotators were not informed of either the project's aim or crisis data, the high inter-coder reliability we obtained alleviates the concern. However, while this validates the labeling given my coding rule, it does not tell us about the reliability of the rule itself. For example, it might be too expansive and inappropriately pool different types of emotions or behaviors.

I first discuss the rationale behind the coding scheme. The four categories of anger expressions—direct and indirect expressions, verbal assaults, and accusations—

are motivated by existing theories and measures of anger expressions. For instance, emotion scholars suggest that the most common angry behavior is what they refer to as "anger-based argumentation," wherein the angry individual verbally communicates his or her displeasure with the perceived mistreatment in quick succession (Averill 1982; Sell et al. 2017). Verbal aggression is another common behavior in anger, which includes cursing, calling others names, and making negative comments about the agents that triggered anger (Deffenbacher et al. 1996). Consequently, my coding scheme is developed with a focus on what an angry leader might say to an adversary. I take this approach because individuals may vary in their ways of expressing anger. Moreover, to my best knowledge, there is no systematic research on particular styles of anger expressions preferred by each leader in my sample or on the relative effectiveness of different expressions. Without such guidance, the current coding scheme has adopted the four categories to capture enough variation in communication styles while maintaining tractability.

Nonetheless, there might still be concerns that the coding scheme inappropriately combines different categories of emotions. Moreover, that individuals have varying levels of sensitivities to others' emotion expressions (Mayer, Caruso and Salovey 1999) might raise questions about whether the statements labeled as angry were actually perceived as such by their targets. The main analysis partly addresses the first issue by including a measure of emotionality derived from the NRC Word-Emotion Association Lexicon (NRC) (Mohammad and Turney 2013). Here, I conduct the same analysis with five different alternative anger measures to deal with the issues more comprehensively. First, I use the NRC lexicon again to create an alternative anger score, which is defined as the proportion of anger words among all words a leader has spoken in a crisis. One important advantage of NRC is that its labeling was crowed-sourced through Amazon's Mechanical Turk. Therefore, NRC emotion words represent laypeople's understanding of

emotionality associated with them. As shown in the "NRC" column of Table 4.5, using this variable as a measure of anger does not alter the main findings. This provides strong evidence that overall, the original measure captures anger as understood by the average English speaker.

Second, I remove each one category of anger expression from the coding scheme to examine the sensitivity of the statistical results to the exclusion of specific categories. Generally, lower sensitivity implies greater robustness of the results, but I do not expect each category to have equal weight. The use of explicit emotion words may make direct expressions the least noisy way to convey anger. In light of the concern under consideration about the noise in the coding scheme, we would be most confident about this category and anticipate that the results will be particularly sensitive to its exclusion. In addition, I expect that the results will be relatively sensitive to the exclusion of the accusation category. Although accusation might be noisier than other expressions as it encompasses a variety of context-specific messages, as discussed above, it is the most common way to verbally communicate anger (Averill 1982). Thus, I contend that excluding accusation will significantly increase the risk of committing type II error by removing much variation in my data.

Table 4.5 describes the estimates of models whose anger measures include "no direct" expressions, "no indirect" expressions, "no insults," and "no accusations." The second column shows that excluding direct expressions weakens the statistical significance of the main findings. This result is consistent with the expectation that direct expressions communicate anger in the least noisy way. The third and fourth columns show that removing indirect expressions and verbal assaults does not change the results, demonstrating the robustness of the results. Examining the fifth column, we see that excluding accusations renders the main results not statistically significant at the conventional level. As expected, this is likely because we lose

significant variation: removing accusations decreases the Challenger's mean anger score by 39% and the Defender's by 43%. While the interaction coefficient for the Defender becomes significant at the 90% level, it nonetheless fails to reach the conventional significance level.

In sum, the results are largely robust across five alternative measures of anger expressions. In particular, the fact that using the crowd-sourced measure of anger as an alternative independent variable does not change the results strongly validates the original measure. The subsequent stress test suggests that my coding of anger expressions is not dependent on the inclusion of indirect expressions and verbal assaults. While this is not the case for direct expressions and accusations, I have provided theoretical justifications for why they might be necessary to capture enough variation in anger expressions. My contention is that excluding them raises the risk of type II error more than it lowers the risk of type I error because they are theoretically more important than the other categories.

	NRC	No direct	No indirect	No insults	No accusations
Anger, A × Threat, A	1498.965***	28.348*	43.975***	54.361***	55.141*
	(488.778)	(16.822)	(16.213)	(20.723)	(31.901)
Anger, A	−73.041*	2.020	0.721	0.164	1.312
	(41.122)	(2.520)	(2.486)	(2.635)	(5.371)
Threat, A	−23.003**	1.279	−2.918	−2.220	−0.987
	(9.481)	(3.821)	(3.599)	(3.602)	(2.698)
Anger, B × Threat, B	185.823	−9.264	−4.310	12.755	−27.461*
	(191.651)	(15.007)	(15.230)	(13.805)	(16.437)
Anger, B	18.880	−0.213	0.097	−1.417	0.247
	(44.520)	(2.177)	(2.219)	(2.238)	(2.271)
Threat, B	−12.718***	−4.265	−5.894	−10.275**	−3.394
	(4.493)	(5.253)	(5.487)	(4.824)	(2.939)
Statement, A	2.453	1.468	1.750	1.539	2.422*
	(1.506)	(1.343)	(1.384)	(1.360)	(1.404)
Statement, B	2.349**	1.316	1.441	2.004	1.078
	(1.088)	(1.219)	(1.132)	(1.223)	(1.184)
Emotionality, A	−6.057	−10.734	−10.645	−8.423	−14.707**
	(9.211)	(7.063)	(7.385)	(7.337)	(7.192)
Emotionality, B	−16.286**	−9.237	−10.599	−12.725*	−6.703
	(7.564)	(7.126)	(7.109)	(7.415)	(6.751)
Word count, A	−0.003	−0.008**	−0.007*	−0.006*	−0.007**
	(0.003)	(0.004)	(0.004)	(0.004)	(0.003)
Word count, B	0.002	0.003*	0.003	0.003	0.003**
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Personalist, A	−0.737	−0.986**	−0.894*	−0.750	−0.985**
	(0.487)	(0.487)	(0.494)	(0.496)	(0.480)
Personalist, B	0.671	0.538	0.590	0.760*	0.408
	(0.454)	(0.459)	(0.481)	(0.455)	(0.470)
Leadership change, A	0.850**	0.851**	0.869*	1.001**	0.736*
	(0.396)	(0.421)	(0.447)	(0.445)	(0.443)
Leadership change, B	0.755	0.384	0.414	0.610	0.541
	(0.495)	(0.566)	(0.576)	(0.552)	(0.584)
Relative power, A	0.743	0.911	0.789	0.740	0.809
	(0.578)	(0.577)	(0.567)	(0.553)	(0.573)
Level of violence	0.156	0.151	0.110	0.074	0.225
	(0.175)	(0.173)	(0.177)	(0.176)	(0.174)
Crisis duration	−0.001	−0.001	−0.001	−0.001	−0.001
	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
Military issue	0.354	0.277	0.299	0.371	0.206
	(0.406)	(0.439)	(0.428)	(0.418)	(0.436)
1|2	−0.163	−0.243	−0.371	−0.337	−0.248
	(0.598)	(0.588)	(0.590)	(0.577)	(0.585)
2|3	2.844***	2.785***	2.674***	2.682***	2.849***
	(0.662)	(0.664)	(0.660)	(0.651)	(0.666)
Num. obs.	179	179	179	179	179

***$p < 0.01$; **$p < 0.05$; *$p < 0.1$. Robust clustered standard errors at the dyad level.

Table 4.5: Interaction effects of anger expressions and threats on crisis outcomes using ordered logit regression

4.5 Exploring Scope Conditions

In the theory chapter, I claimed that anger will be more useful for challengers than for defenders. Defenders' threats may carry some inherent credibility due to their perceived or real willingness to take risk to protect the status quo (Art and Greenhill 2018), and their anger expressions might be anticipated by challengers as a natural reaction to a challenge to national interests. If defenders' threats are already likely to be credible and their anger is expected, then their anger expressions will make only marginal impact on credibility. Challengers' anger expressions will be more useful because challengers make relatively less credible threats and because their anger is less anticipated by defenders. The statistical analysis described in this chapter provides the results that are aligned with this scope condition. In this section, I empirically investigate the plausibility of my claims about the scope.

First, are threats from defenders more credible than threats from challengers? Challengers prevail 17% of the time and back down 29% of the time in the bilateral crises they are involved in. These raw statistics point to a defender's advantage in bargaining, but is this because defenders make more threats than challengers? To assess the independent effects of threats, I estimate an ordinal logit model similar to the one reported in the main manuscript but without anger scores and emotional measures. The results depicted in Table 4.6 suggest that while both challengers' and defenders' threats exert a significant influence on winning outcomes, defenders' threats are more effective. A 0.01 increase in the challenger's anger score is associated with a 0.7 percentage point increase in its probability of winning, whereas the same increase in the defender's anger leads to a 0.89 percentage point decrease in the challenger's probability of winning.

	Model 1
Threat, A	5.723 (2.389)**
Threat, B	−7.264 (2.124)***
Statement, A	−0.007 (0.482)
Statement, B	0.164 (0.450)
Word count, A	−0.002 (0.003)
Word count, B	0.002 (0.001)
Personalist, A	−0.498 (0.451)
Personalist, B	0.575 (0.428)
Leadership change, A	0.666 (0.442)
Leadership change, B	0.655 (0.469)
Relative power, A	0.806 (0.566)
Level of violence	0.134 (0.173)
Crisis duration	−0.001 (0.001)
Military issue	0.291 (0.402)
1|2	−0.095 (0.571)
2|3	2.736 (0.630)***
Num. obs.	179

*** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$. Robust clustered standard errors at the dyad level.

Table 4.6: Independent effects of threats on crisis outcomes

Second, is defenders' anger anticipated by challengers but not the other way around? While it is difficult to answer this question directly with the available data, we can look for indirect evidence by considering the predictability of anger. If defenders exhibit behavior typical of angry individuals in the real world more often than challengers, this would suggest that defenders' anger might be more likely to be anticipated. We should first expect defenders to express more intense anger than challengers. A simple comparisons of means shows that on average, defenders' anger scores (M = 0.10, SD = 0.17) are about 65% higher than challengers' (M = 0.06, SD = 0.13), a difference that is substantially large and statistically significant ($p < 0.01$).

In addition, we would expect defenders to express greater anger at the outset of a crisis than later stages. The intuition here is that if anger can be anticipated because it is an automatic reaction to a challenge, then we should see that the defender meets a challenge with anger immediately. However, if we instead find

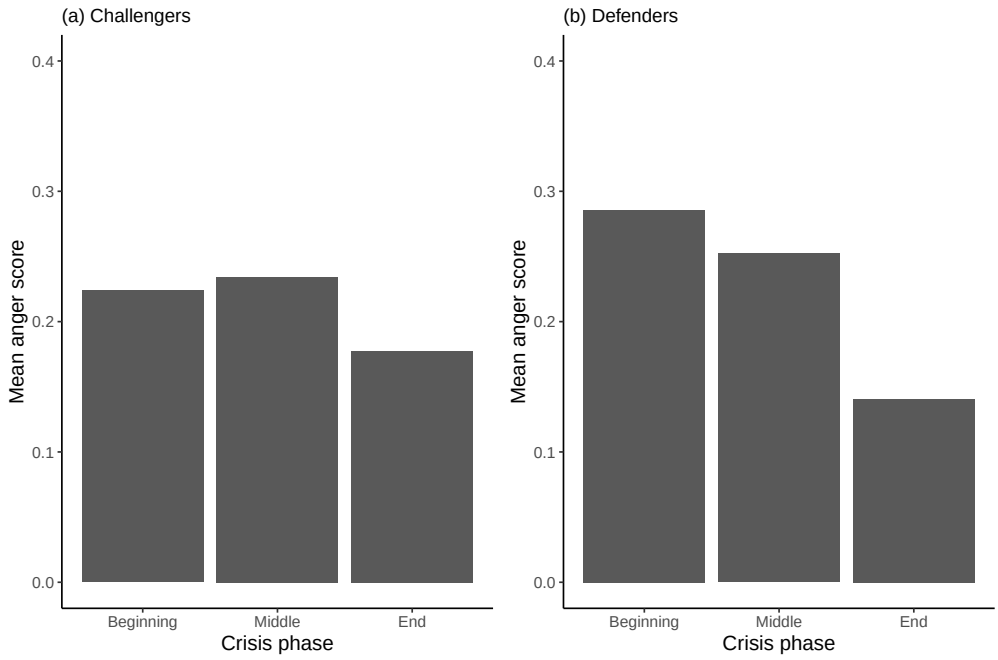

Figure 4.2: Distribution of anger scores for challenger and defenders by phase

more anger expressions in the later stage of a crisis, then at least two explanations are possible. First, anger might not be an automatic reaction but rather a consequence of negative interactions throughout the crisis, and challengers would have no reason to anticipate anger from defenders in advance. Second, even if we maintain the assumption that challengers expect defenders' anger, a lack of immediate angry responses could lead them to infer that defenders are unresolved. Under this condition, the defender's anger expressions are informative. In either case, we would need to reconsider the null result for the defender's anger as lack of support for my hypothesis. Conversely, if we find that defenders express greater anger at the beginning than at the end of a crisis, this will allow me to rule out these alternative explanations, strengthening the plausibility of my scope condition.

Examining the distributions of anger scores in Figure 4.2, we see different patterns of anger expressions from challengers and defenders. Taking the same approach I used in the "Addressing Confounding" section to generate Table 4.3, I

	Challenger's anger	Defender's anger
(Intercept)	0.109 (0.062)*	0.238 (0.029)***
Middle	−0.006 (0.089)	−0.057 (0.045)
End	−0.124 (0.082)	−0.271 (0.053)***
Total	254	477
Left-censored	113	193
Uncensored	132	274
Right-censored	9	10

*** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$. Robust clustered standard errors at the crisis level.

Table 4.7: Tobit models predicting challengers' and defenders' mean anger scores by crisis phase

subset crises that involve public statements and divided each crisis' duration into three equal chunks (beginning, middle, and end). Panel (a) shows that challengers express similar levels of anger throughout the crisis. However, in Panel (b), defenders seem to express the highest level of anger at the crisis's onset, and tone down their anger moving toward its termination. Table 4.7 summarizes the results of a formal test of this pattern, predicting anger scores by crisis phase using tobit regression.[8] The results confirm that while challengers' level of anger indeed does not differ by phase, defenders communicate significantly higher anger at the beginning than at the end of the crisis, although the difference between the Beginning and Middle categories is not significant. Together, these supplementary analyses provide a solid empirical basis for the scope condition for my theory.

[8]The references category is Beginning.

Chapter 5

Experimental Evidence of the Effect of Anger on Beliefs about Resolve

The results of the observational analysis shed light on the role of anger statements in the context of international crisis. By utilizing the cross-national measure of elite anger expressions and threats, the observational analysis overcomes the major limitation of emotions research in international relations that there is lack of historical data on leaders' emotions. As such, it offers generalizable evidence that anger expressions increase the credibility of threats across time and space. Yet, we would want to supplement these results for two reasons. First, while the use of observational analysis has allowed us to find broad patterns regarding anger expressions, it simultaneously limits our ability to draw causal inference because anger expressions and threats do not arise randomly. Although I carefully address confounding through a series of supplementary analyses, they only show that the results are robust to the specific issues the data allow me to tackle. They do not support making causal claims about anger expressions. The second reason concerns the unit of analysis. My theory makes predictions about how angry expressions influence targets' beliefs about credibility. However, the observational

analysis results cannot directly speak to the perceptions of the targets at the individual level. Similarly, the macro-level analysis is unable to offer insight into the mechanisms through which anger expressions operate. To establish a more direct, causally identified link between anger expressions and signaling resolve and to provide evidence for the causal mechanism, I turn to a randomized experiment.

5.1 Experimental Design

5.1.1 Sample

In May 2023, I fielded a pre-registered online survey experiment on a sample of U.S. citizens recruited through a survey platform Lucid. A total of 1,672 respondents initially entered the study; I terminated from the survey 89 respondents who did not consent to participate and additional 526 who failed any of the three easy pre-treatment attention checks. The remaining 1,012 respondents were, on average, 49.7 years old and highly educated, with 47.5% having at least a bachelor's degree. About 75.1% were Caucasian, 48.6% identified as male, 34.7% viewed themselves as Republicans, and 44.6% as Democrats. The sample's demographic information, foreign policy attitudes, and political ideology are summarized in Table 5.1 and Figure 5.1.

5.1.2 Design

In order to test whether anger expressions influence how threats are perceived, the experiment needs to use a crisis scenario in which respondents receive a threat and an anger expression from an adversary. To this end, I adopt a modified version of the vignette and design developed by Yarhi-Milo, Kertzer and Renshon

	Percent Sample
Male	48.60
White	75.10
University education	47.50
Republican	34.70
Democrat	44.60
Independent and other	20.60
Income less than $30,000	34.90
Income $30-60,000	27.10
Income $60-100,000	22.90
Income over $100,000	15.10
Age 18-24	10.30
Age 25-34	19.10
Age 35-44	19.20
Age 45-54	15.20
Age 55+	36.30

Table 5.1: Experiment 1 Sample Demographics

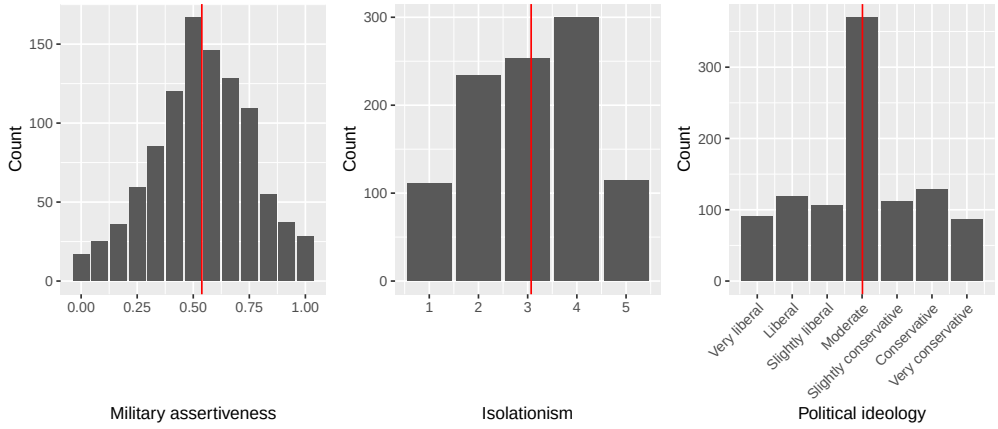

Figure 5.1: Foreign policy attitudes and ideology of the sample

(2018), who find that both the Israeli elites and public update their beliefs about an adversary's resolve in response to its public threat or mobilization. Building on their design offers at least two advantages. First, it provides replication benefits. Finding similar results on a U.S. sample in a different context will increase our confidence in the findings about public threats, contributing to the accumulation of causal evidence regarding the effects of costly signals. Second, more importantly for my purposes, it allows me to directly speak to how the findings of the current

experiment relate to past work. If we find that adding anger expressions alters the effectiveness of threats identified by Yarhi-Milo, Kertzer and Renshon (2018), it will provide support for my claim that anger is a key mechanism for threat credibility.

The experiment has two noteworthy design features. First, it uses what Clifford, Sheagley and Piston (2021) call a "pre-post design", in which outcome variables are measured both before and after random assignment into treatments. By controlling for the pre-treatment measure of the outcome variable, the pre-post design yields a very precise estimate of the main treatment effect, with an additional benefit of higher statistical power. In addition, the within-subject portion of the experiment lets me analyze how respondents update their beliefs to further unpack the treatment effect. The second feature is the use of "implicit mediation" analysis, which "adds and subtracts features of the treatment in ways that implicate some mediators and not others" (Bullock and Green 2021). The most common approach to mediation analysis has been measuring the mediating variable and using regression to understand the causal pathway. However, this method relies on an assumption that there are no unmeasured confounders for the relationship between the mediator and the outcome (Acharya, Blackwell and Sen 2018). In my setting, the assumption is easily violated by, for example, respondents' dispositions that influence both their perceptions of an adversary's anger and their beliefs about its resolve. Failing to control for all of these confounders will violate the assumption and render the mediation results unsustainable. Instead, by exogenously varying anger expressions in a statement within which threats are embedded, the implicit mediation design allows me to isolate the effect of anger on how threats are perceived. The consort diagram of the experimental design is depicted in Figure 5.2.

After completing a pre-treatment questionnaire measuring foreign policy at-

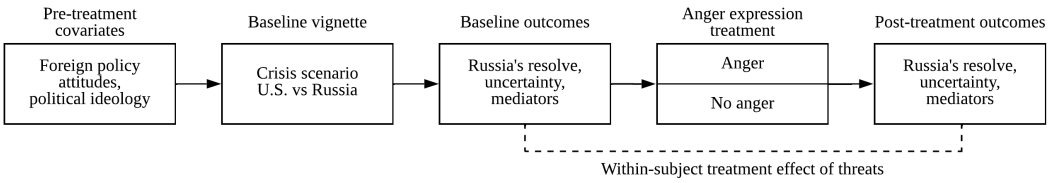

Figure 5.2: Survey experiment design

titudes, respondents read the following introduction to the vignette: "We are going to describe a situation our country has faced many times in the past and will probably face again, and then ask your opinions about it." Respondents are then presented with a hypothetical crisis between the U.S. and Russia that takes place in 2040. The crisis involves a collision between a U.S. shipping vessel and a Russian ship, resulting in injuries on both sides and "a tense standoff at sea," as both ships are carrying sensitive military technology. For the pre-treatment phase of the experiment, respondents are also informed that both countries' leaders have yet to make any statement regarding the dispute. Unlike the original vignette that uses a hypothetical adversary whose regime type is varied between a democracy and a dictatorship (Yarhi-Milo, Kertzer and Renshon 2018), the crisis adversary is Russia in my version. I chose Russia over other historical U.S. adversaries, such as Iran, to keep the power balance between the U.S. and the adversary relatively similar. To hold regime type and relative power constant and maintain the realism of the scenario, the vignette describes Russia as a nondemocracy with a strong military. After reading the scenario, respondents report their pre-treatment beliefs about Russia.

Respondents are then randomly assigned into the "anger" condition or the "no anger" condition. In both conditions, a male Russian leader threatens to use force if the U.S. does not back down, holding the leader's gender and threat constant. I fix gender because the effect of anger might vary based on the leader's gender

due to gender stereotypes of anger (Plant et al. 2000).[1] In the anger condition, the Russian leader "yells in outrage" while issuing a threat and adds that he will "make it clear that we are angered by how America treated us." On the other hand, in the no anger condition, the leader issues the threat "in a calm, measured voice" without any additional remark about his emotion. Explicit information about the leader's emotions is provided to ensure a clear contrast between the two conditions. After the treatments, respondents completed a post-treatment questionnaire comprising the same set of questions as the pre-treatment outcomes and additional exploratory variables, which will be described in the next section.

5.1.3 Outcomes

The main outcome of the experiment is the belief about resolve. Following Yarhi-Milo, Kertzer and Renshon (2018), it is measured by asking respondents to provide an estimate of the probability that Russia will use force against the U.S. to resolve the dispute. The probability of using force reflects the concept of willingness to fight in rationalist models of crisis bargaining (Fearon 1995). Alternatively, scholars often define resolve more narrowly as willingness to bear costs (Kertzer 2016; Ramsay 2017). Thus, I additionally measure beliefs about the Russian leader's sensitivity to military casualties. Granted, international conflict has enormous social, economic, and environmental ramifications (Modell and Haggerty 1991; Smith 2014; Rawtani et al. 2022). I nonetheless focus on military casualties because they are one of the most politically salient type of cost that affects both decision makers' and the public's attitudes towards military conflict (Gartner 2008; Gelpi, Feaver and Reifler 2009). Military casualties may also be normatively more important than other costs due to the perceived sanctity of human lives (Tetlock 2003).

[1] Nonetheless, if respondents are not provided with any information about the leader's gender, they are likely to assume a male leader because Russia has never had a female leader.

The variable is measured by asking respondents to indicate their "best estimate about how many military casualties Russia's leader is willing to accept to win this dispute" (Possible choices: zero, 3, 30, 300, 3,000, 30,000, and any number of casualties).

To explore the broad impact of anger expressions, I measure respondents' perceptions about several outcomes: stakes, honor concerns, and public disapproval. First, people are more likely to become angry over issues they care about deeply than over issues they care about less. This means that the leader's anger might alter perceptions about the interests at stake in the crisis. Given the setup of the vignette, it is natural that what is at stake is the sensitive military technology the two countries' vessels are carrying. Therefore, I ask respondents to what extent they believe that "the military technology the Russian ship was carrying is valuable for Russia." Second, anger expressions might also affect perceptions of immaterial concerns such as national honor. To investigate this possibility, respondents indicate whether they believe the leader "intends to maintain Russia's national honor." The third variable measures respondents' beliefs about whether "the Russian leader will face strong public disapproval if Russia does not use force." The term *public disapproval* here captures a bundle of costs for the leader. It could be audience costs of backing down from a threat (Tomz 2007*a*), backlash against tarnishing the country's international reputation (Brutger and Kertzer 2018), or disapproval of perceived incompetence (Nomikos and Sambanis 2019). Respondents are asked to answer these questions on a 5-point scale, where 1 means "Strongly disagree" and 5 means "Strongly agree."

Lastly, I measure three variables to test additional implications of my theory. First, respondents are asked to indicate their level of certainty regarding their estimates of the probability of Russia's using force, on a scale from 0% to 100%. In Chapter 2, I argued that anger expressions might trigger fear as well

as counter-anger in targets in order to address a concern that the counter-anger might ultimately make threats less effective despite increased credibility (Powers and Altman 2023). Because fear decreases risk attitudes and causes a flight response, it will offset some of the effects of anger on decision-making (Lerner and Keltner 2001). As a first step toward empirically testing this idea, I solicit respondents' self-reported hostility and fear using the PANAS-X scale (Watson and Clark 1994). Finally, I measure perceptions of Russia's military power to ensure that anger expressions do not operate through changing beliefs about power. These variables are measured post-treatment only due to concerns about carry-over effects (Clifford, Sheagley and Piston 2021). Asking about certainty and emotions pre-treatment might unintentionally prime respondents to be more self-aware of their emotions or more attentive when answering the post-treatment questions, potentially masking the true treatment effect.

5.2 Results

5.2.1 Manipulation and Attention Checks

First of all, manipulation checks validate the treatments. I included two manipulation checks after the anger treatment. The first check item asked respondents whether the Russian leader in the scenario threatened to use force against the U.S. or signed a trade agreement with the U.S. Across the two experimental conditions, 92.6% of the respondents correctly answered the question, suggesting that the within-subject threat treatment worked as intended. The second check asked how angry the Russian leader was on a scale of 1-5, where 1 means "Not angry at all" and 5 is "Extremely angry." A t-test suggests the anger manipulation was successful: respondents in the anger condition ($M = 4.40$, $SD = 0.90$) believed that the Russian leader was significantly angrier ($t(938.5) = 25.8$, $p < 0.001$) than those

in the no anger condition (M = 2.64, SD = 1.25).

The survey includes three pre-treatment attention check questions, drawn from Berinsky et al. (2021). They impose very low cognitive costs and are design to screen out extremely low attention respondents.[2] I also include one post-treatment attention check at the end of the survey to gauge how many respondents remain attentive. The attention check showed that 94.7% of the respondents who passed the initial checks retained their level of attention throughout the survey. Those who failed the post-treatment checks are included in the analysis because dropping them can cause post-treatment bias (Montgomery, Nyhan and Torres 2018).

5.2.2 Anger Expressions Increase Beliefs about Resolve

The results illustrated in Figure 5.3 compare respondents' perceptions of Russia's resolve when the threatening leader is angry versus not angry. Across three outcomes, the results provide striking evidence supporting leader anger as an important source of credibility, confirming Hypothesis 2. Examining Panel (a), when the Russian leader did not express anger, respondents estimated a 50.6% probability that Russia would use force. Manipulating information about the leader's anger expression increased this estimate by 14.6 percentage points ($p < 0.01$). Despite a public threat from the Russian leader, respondents essentially estimated an even chance of following through or backing down when he did not appear angry. After observing the leader's anger, however, respondents reported a significantly higher probability of following through.

If my theory is correct in claiming that anger has informational value, we

[2]The first two questions are embedded in a question matrix measuring foreign policy attitudes. Respondents answer the following two questions on a scale of 1 to 5, in which 1 is "Strongly disagree" and 5 is "Strongly agree": "World War I came after World War II," and "Please click Neither agree nor disagree." The third question is attached to the introduction to the vignette: "Do you agree to read the details carefully and give thoughtful answers?" Those who chose "no" instead of "Yes" failed the check.

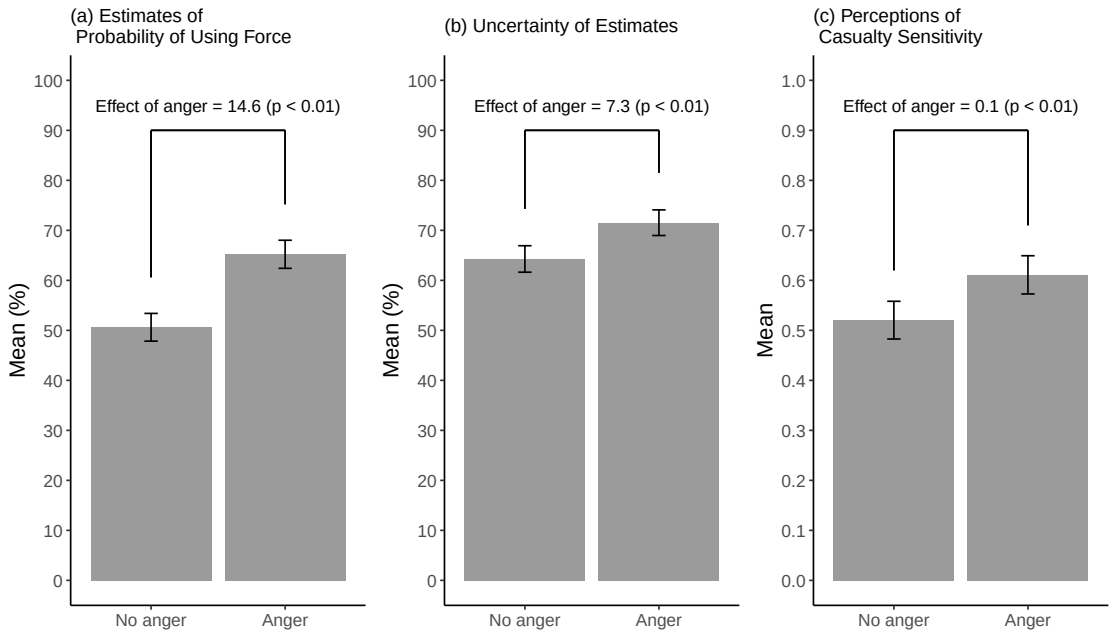

Figure 5.3: Main treatment effects
The plot displays the treatment effects on (a) estimates of the probability of using force, (b) uncertainty of the estimates, and (c) perceptions of casualty sensitivity. The uncertainty estimates are only relevant to the probability of using force. Perceptions of casualty sensitivity are measured on a discrete scale from 1-7 and normalized to a 0-1 range. All estimates are rounded to the nearest tenth. The p values are adjusted for multiple comparisons through the Benjamini and Hochberg method.

would expect the anger treatment to not only increase respondents' estimates of resolve but also increase their certainty about these estimates. In this light, the estimated 50.6% chance of using force in the no-anger condition in Panel (a) might indicate high uncertainty about the outcome. Panel (b) supports these expectations, confirming Hypothesis 4. Respondents in the anger condition were about 7.2 percentage points more certain about their estimates of Russia's probability of using force than those in the no anger condition ($p < 0.01$). However, we should be cautious about concluding that this effect is unique to anger. The analysis compares anger with a lack of anger or calmness rather than with a different emotion. The current design cannot determine whether this effect is driven by the theorized mechanism of anger or any emotion can increase certainty simply by

	Probability of Using Force	Casualty Sensitivity	Uncertainty
Intercept	11.001 (1.529)***	0.183 (0.021)***	64.278 (1.315)***
Anger	15.071 (1.383)***	0.088 (0.022)***	7.254 (1.877)***
Resolve, T1	0.736 (0.022)***		
Casualties, T1		0.603 (0.025)***	
R^2	0.550	0.370	0.015
Num. obs.	1012	1012	1012

***$p < 0.01$; **$p < 0.05$; *$p < 0.1$.

Table 5.2: Treatment effects on dependent variables
OLS coefficients are shown. P-values are adjusted for multiple comparisons of three hypotheses using the Benjamini and Hochberg method. While this differs from the pre-registered analysis that compares two hypotheses, this is more robust. The casualties variable is rescaled to 0-1. Per my preanalysis plan, the models for casualties and resolve control for the pre-treatment measures. The other model is bivariate.

providing an additional informative cue.

Next, focusing on perceptions of casualty sensitivity as an alternative measure of resolve, Panel (c) similarly demonstrates a substantial impact of anger on beliefs about resolve ($p < 0.01$). To provide some context to illustrate its effect, the proportion of respondents who indicated that the Russian leader would accept zero military casualties decreased from 32.8% to 26.9%, and those who indicated that the leader would tolerate 3,000 casualties, which is roughly the number of U.S. fatalities in the Iraq War, increased from 48.5% to 59.1%. In effect, expressions of anger led the majority of respondents to believe that the leader was willing to risk a war. The full estimates are summarized in Table 5.2. The p-values are adjusted for multiple comparisons of three hypotheses using the Benjamini and Hochberg method.

Finally, Figure 5.4 illustrates the broad effects of the treatment on other outcomes that we might or might not expect to be plausibly affected by anger expressions in real-world scenarios. First, the treatment increased perceptions of the stakes of the crisis and the public disapproval the Russian leader would face

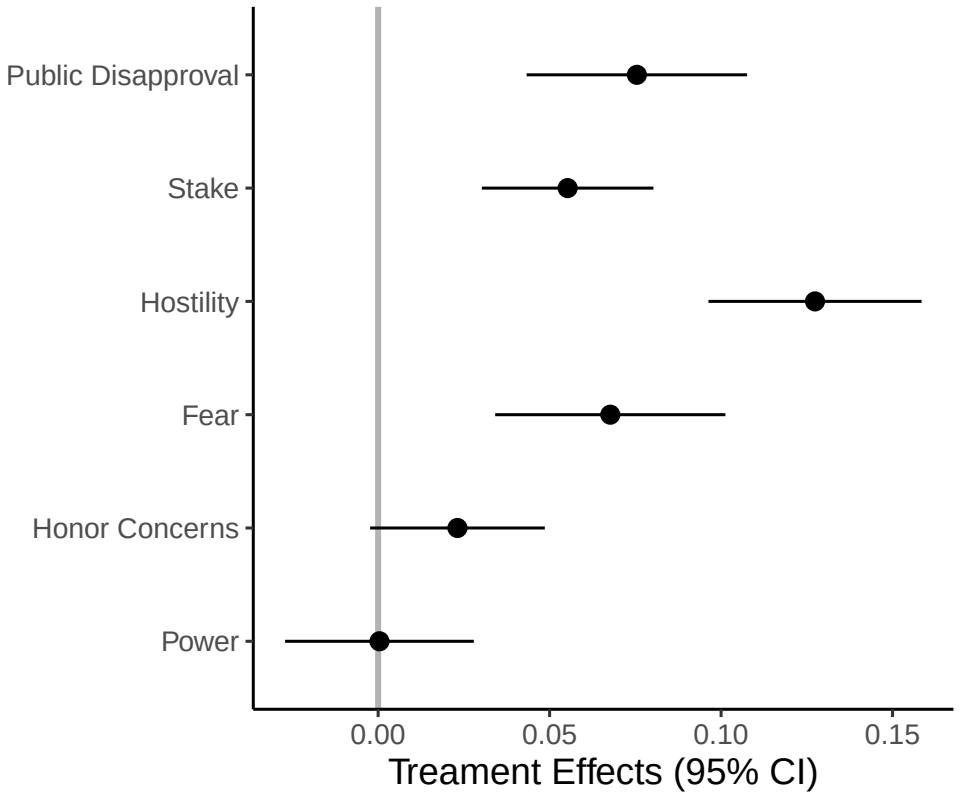

Figure 5.4: Main treatment effects on other exploratory outcomes
The variables are rescaled to a 0-1 range.

for backing down, but did not change those of Russia's honor concerns. Anger may signal high interests but may not automatically raise the salience of national honor. Second, the null effect on power suggests that anger expressions do not influence international crises by changing perceptions about relative power. Third, the anger treatment caused fear as well as hostility in targets. Although the effect size on hostility is larger, fear might offset some of the urge to resist caused by hostility (Powers and Altman 2023).

5.2.3 Anger Expressions Lead to Differential Updating

The main effects provide compelling evidence that anger expressions increase the credibility of threats. Another approach to analyzing the results is to examine

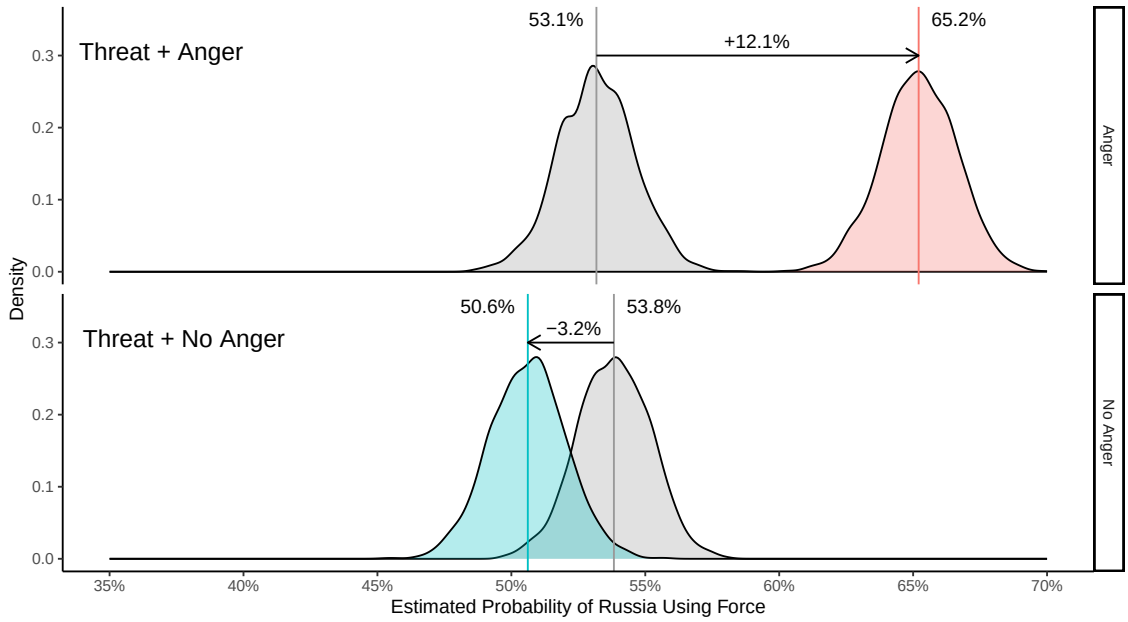

Figure 5.5: Within-subject effect of anger expressions
The plot depicts the effects of anger on estimates of the probability that Russia will use force. The plot presents the bootstrapped distributions of the treatment effects. The baseline measures are color-coded as gray. The post-treatment outcomes for the anger and the no anger conditions are color-coded as red and blue, respectively. All estimates are rounded to the nearest tenth

how manipulating information about anger leads respondents to update their beliefs about resolve differentially. More specifically, as Hypothesis 3 states, we would expect to see that respondents update their beliefs about Russia's resolve to a greater extent when they receive threats from an angry leader, relative to when they receive threats from a calm leader. Figure 5.5 describes the distributions of the mean pre- and post-treatment estimates of the probability of Russia using force in the anger condition (the top panel) and the no anger condition (the bottom panel). The gray distributions represent the pre-treatment measures, and the red and blue represent the post-treatment measures for the anger and no anger conditions, respectively. Subtracting the pre-treatment measure of resolve from the post-treatment measure within each group gives the within-subject treatment effects of a threat conditional on anger expressions.

The results lend support for my expectation. The top panel of Figure 5.5 shows that after receiving a threat from the angry leader, respondents increased their estimates of the probability that Russia will use force from 53.1% to 65.2% ($p < 0.01$). The bottom panel, on the other hand, shows that the non-angry leader's threat *decreased* perceived resolve by 3.2 percentage points ($p < 0.01$). These results strongly support the contention that anger is a key mechanism for threat credibility, suggesting that threats devoid of anger might backfire. The results are also consistent with the observation from Study 1 that threats have little effect on crisis outcomes when they are not conveyed in conjunction with anger.

Nonetheless, the negative updating due to calm threats is surprising in light of existing evidence. For instance, Yarhi-Milo, Kertzer and Renshon (2018) find that the Israeli elites and members of the public reported a 8.1% chance and a 22% chance of using force as a result of receiving public threats. Kertzer, Renshon and Yarhi-Milo (2021) find that the U.S. public views countries that issue public threats as 6.4% more likely to stand firm in crises they are involved. The design of the current experiment does not permit a conclusive investigation into the reason for the negative effects. Yet I conjecture that in the earlier experiments, respondents viewed the leader making threats as angry. People may anticipate crises to evoke anger; without any information about the leader's emotion, respondents may still "fill in the blanks" using their intuition (Brutger et al. 2023). The backfiring effect of public threats identified in my experiment might occur outside of a controlled experimental setting. In real-world crises, observers likely have access to information about the emotions of leaders making public appearances, either by directly observing them or due to greater media coverage of emotional statements (Maier and Nai 2020).

5.3 Alternative Causal Mediation Analysis

The analysis thus far has focused on the implicit mediation design and discussed perceptions of casualty sensitivity as an alternative measure of resolve. In the theory chapter, I argued that angry leaders can project an image of resolve by appearing relatively insensitive to the cost of fighting. In this light, another way to frame my theory is to posit casualty insensitivity as a causal pathway through which anger expressions influence targets' beliefs about resolve. I can then evaluate my theory by analyzing the extent to which perceptions of casualty sensitivity mediates the effect of the treatment effects of anger expressions, utilizing the causal mediation analysis method proposed by (Imai, Keele and Yamamoto 2010). The main analysis used in this chapter uses the implicit mediation design because it obviates the need for the strong sequential ignorability assumption to sustain the mediation results (Imai, Keele and Yamamoto 2010). However, if we find similar results in an alternative approach, it will increase our confidence in the findings. Furthermore, the causal mediation approach allows me to explore other potential mechanisms of anger expressions that are not explicitly accounted for in my theory, possibly suggesting directions for further theorization about the role of anger in international crises.

The first step in this framework is to demonstrate the causal effects of the anger treatment on the theorized mediator. Panel (a) in Figure 5.6 shows that anger expressions have positive and statistically significant causal effects on perceptions of cost insensitivity ($p < 0.01$).[3] U.S. respondents produced higher estimates of their adversary's casualty sensitivity when they observed an expression of anger.

I now turn to the causal mediation analysis, investigating the extent to which the effect of the anger treatment was mediated by perceptions of low sensitivity

[3]This estimate is the same as the one shown in Figure 5.3 in the previous section. Panel (a) shows the treatment effects on cost insensitivity and other variables, effectively summarizing Panel (c) of Figure 5.3 and Figure 5.4.

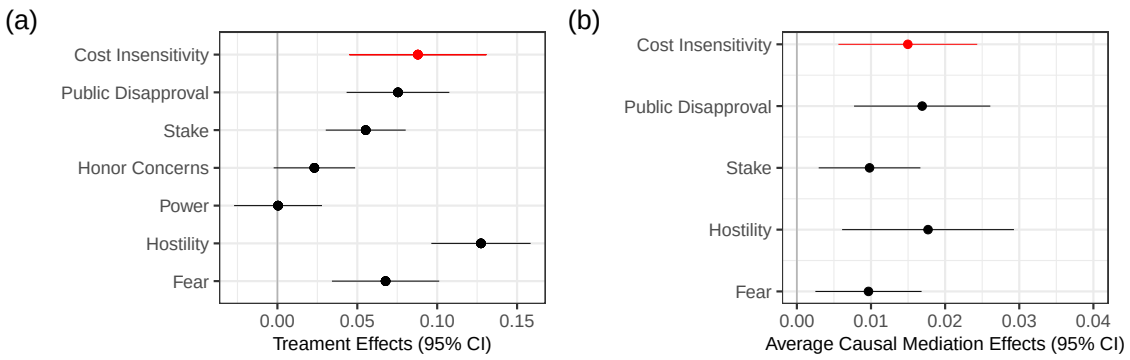

Figure 5.6: Causal mediation analysis for the anger treatment
Panel (a) depicts the treatment effects of the anger treatment on the mediators estimated from bivariate OLS regression models. The model for the cost insensitivity mediator includes its pre-treatment measure as a covariate to improve precision. Panel (b) describes the average causal mediation effects. Only mediators that are causally affected by the treatment are included. Error bars are 95% confidence intervals. Red points indicate the theorized mediator.

to the cost of fighting. I also examine five additional mediators—perceptions of public disapproval for backing down, stake, honor concerns, and respondents' hostility and fear—to explore other mechanisms. First, confirming Hypothesis 5, I find evidence that the cost insensitivity mechanism mediates the effect of anger expressions. Panel (b) of Figure 5.6 describes the average causal mediation effects (ACMEs) and 95% bootstrapped confidence intervals. The ACME is a measure of the proportion of the main effect of the anger treatment on perceived resolve that is caused by the treatment's effect on each of the mediators. The models used in estimating the ACMEs account for causal dependence between alternative mediators while controlling for a number of pre-treatment covariates, including age, gender, ideology, education, hawkishness, and party ID. Perceptions of cost insensitivity accounts for 10.7% of the total effect of the anger treatment on estimates of resolve. This analysis provides support for the claim that cost insensitivity plays a substantial role in signaling resolve.

Next, the results suggest that anger expressions move other potential mediators. As shown in Panel (a) of Figure 5.6, anger expressions increased the beliefs

that the leader would face strong public disapproval for backing down and had a high stake in the crisis, as well as respondents' own emotions. Panel (b) suggests that these beliefs in turn mediated the effects of anger expressions. The belief about honor concerns was not affected. While the analyses provide suggestive evidence for other independent mechanisms, this suggests that the causal mediation estimates could be biased. To address this issue, the mediation models reported here statistically account for dependence between mechanisms (Tingley et al. 2014). However, confounding from treatment-mediator interactions could still bias the causal mediation estimates. While the more precise evidence for causal mechanisms requires a design that manipulates cost perceptions, I conduct sensitivity analysis to examine the robustness of the results.

5.3.1 Sensitivity Analyses

I used the mediation models that account for dependence between mechanisms (Tingley et al. 2014). However, these models still require that there is no interaction between the observed mediators and the treatment. In this section, I discuss results from sensitivity analyses that evaluate the robustness of the ACME estimates when that assumption is violated.

The sensitivity analyses use two different parameters: σ and $\widetilde{R}^2$. In the context of this study, the σ values represent the levels of heterogeneity in the interaction between the mediators and estimates of resolve. The $\widetilde{R}^2$ indicates the proportion of total variance of estimates of resolve that could be explained by the treatment-mediator interaction (Imai and Yamamoto 2013). Regardless of which parameter is used, higher values indicate greater robustness to the violation of the assumption of no treatment-mediator interaction.

The top row in Table 5.3 displays the outcomes of the sensitivity analyses with respect to the main mediator of interest, the leader's cost insensitivity. The maxi-

mum value of σ, 0.505 presents the greatest possible violation of the no-interaction assumption. Examining the σ value, the results suggest that the lower bound of ACME crosses zero when $\sigma > 0.051$, which is 9.9% of its largest possible value. Using an alternative parameterization, $\widetilde{R}^2$, the results suggest that ACME estimate becomes indistinguishable from zero when $\widetilde{R}^2$, > 0.007, or when the treatment-mediator interaction explains 0.7% of the variance in estimates of resolve. Together, these estimates suggest that my findings about the mediating effects of cost sensitivity might be relatively vulnerable in the presence of the treatment-mediator interaction. The results thus imply that a more definitive conclusion about the cost insensitivity mechanism would require a study that directly manipulates the mediator rather than observationally measuring it.

	ACME	CI.Low	CI.High	σ	Max.σ	$\widetilde{R}^2$	Max.$\widetilde{R}^2$
Cost insensitivity	0.015	0.006	0.024	0.051	0.505	0.007	0.654
Public disapproval	0.017	0.008	0.026	0.081	0.573	0.013	0.654
Stake	0.010	0.003	0.017	0.067	0.474	0.013	0.654
Hostility	0.018	0.006	0.029	0.079	0.787	0.007	0.654
Fear	0.010	0.002	0.017	0.081	0.806	0.007	0.654

Table 5.3: Sensitivity analyses for anger expressions through the cost insensitivity mediator

It is worth discussing the results of the sensitivity analysis with respect to other mechanisms. For the public disapproval mediator, the ACME lower bound crosses zero when $\sigma > 0.081$. For the stake mechanism, the ACME lower bound crosses zero when $\sigma > 0.067$. For both variables, these values represent 14.1% of their largest possible values. These results have interesting implications. First, the fact that perceptions of public disapproval appears relatively robust to the no-interaction assumption suggests the importance of this mechanism and calls for a more serious theorization about how anger expressions operate through it. Second, while my theory focuses on cost insensitivity, the relative robustness of the stake variable suggests that anger expressions might also work by changing targets'

perceptions of what the leader believes is at stake. This mechanism is plausible given the information asymmetry in crisis situations. In the experiment, for example, the respondents were only informed that the Russian ship was carrying military technology, not how important the technology is to Russia. Intense anger from the leader may convey information about the leader's valuation of the technology. Next, for hostility and fear, the ACME estimates become zero when σ equals 0.079 and 0.081, respectively, or about 10% of their highest possible values. The parameterization using $\widetilde{R}^2$ produces results identical to that of the cost insensitivity mediator. These suggest an interesting possibility that targets' own emotions might contribute to the credibility of threats, such that having "strong feelings" about a threat increases the perceived resolve of the threatening adversary.

5.4 Addressing External Validity Concerns

The experiment has provided some of the first causal evidence for the role of anger expressions in crisis bargaining. However, one important consideration is external validity, the extent to which the causal findings can be generalized to other contexts. One purpose of the observational analysis in the previous chapter is to address this concern preemptively by showing that anger statements matter across time and space. Nevertheless, it is still important to discuss the external validity of the experimental results themselves. Two questions about external validity are particularly relevant to my experiment: populations and contexts (Egami and Hartman 2023). First, do the findings hold on a sample of foreign policy elites? That is, my theory concerns leaders, but it is tested on a public sample, leaving a gap between the target population and the sample. Second, do the findings generalize to non-U.S. contexts? Because my theory is about interstate

interactions, it is important to understand whether and how anger expressions affect perceptions of resolve in countries beyond the U.S. I will discuss each issue in turn.

First, recent evidence suggests that elites and ordinary citizens may respond very similarly to experimental treatments. Through a meta-analysis of 162 public-elite paired experiments, Kertzer (2022) finds that despite differences in individual characteristics, "the treatment effects recovered in the elite samples did not significantly differ in magnitude from those recovered from mass samples 88% of the time," suggesting that political scientists might "overstate the magnitude of elite-public gaps in decision-making." Kertzer additionally argues that much of the elite-public gap can be attributed to compositional differences in the sample, such as oversampling of highly educated, wealthy men, rather than domain-specific elite expertise or experience. This is not to imply that we can simply assume external validity to elites. Rather, these findings provides us with a useful framework to investigate the issue empirically. If "eliteness" is determined by a set of individual-level moderators, we can explore heterogeneous treatment effects with respect to elite-like traits to understand whether they alter the magnitude or the sign of the treatment effect of anger expressions.

The second external validity issue to tackle is the context. A recent study by Bassan-Nygate et al. (2023) offers evidence that experimental findings about international relations theories can generalize across multiple countries. After successfully replicating several benchmark international relations experiments, including those on audience cost and democratic peace, in seven countries, Bassan-Nygate et al. (2023) draw conclusions similar to those of Kertzer (2022). While the U.S. sample does differ from other populations in terms of foreign policy attitudes, these differences do not alter experimental findings due to low treatment effect heterogeneity. That is, the distribution of treatment effects in the U.S. sample

does not appear to differ significantly from the distribution in non-U.S. samples (Coppock 2019). Bassan-Nygate et al. (2023) also find that to the extent that there is treatment heterogeneity, "it tended to shape the magnitude rather than the direction of effects." I expect this pattern to apply in my setting: Social psychological studies indicate that the expression and perception of anger are commonly understood across different cultures (Hupka, Lenton and Hutchison 1999; Laukka and Elfenbein 2021), suggesting low treatment heterogeneity of anger expressions.

The main lesson from the two studies is that, as Bassan-Nygate et al. (2023) put it, "researchers should theoretically and empirically interrogate the extent to which their effects are heterogeneous ... in which case they should consider how samples in other contexts might differ along relevant covariates." Of course, American citizens will differ from American elites and other countries' citizens on multiple dimensions. The crucial question, however, is not simply whether they are different. Rather, we should ask whether individual traits that influence decision making and interpretations of others' emotional expressions differ across these samples, and to what extent the differences moderate the treatment effects.

To this end, I examine whether the treatment effect of anger expressions is moderated by demographic factors and foreign policy dispositions. Following Mattes and Weeks (2022), I first interact the anger treatment with four proxies of political elites—male, high income, university education, and age between 40 and 65. Second, I conduct the same analysis for race, hawkishness, isolationism, party ID, and ideology to represent varying levels of political and foreign policy attitudes. Some of these variables, such as party ID, is an imperfect approximation because party systems vary significantly across countries. However, it may capture the left-right divide given that political polarization is a worldwide phenomenon (Carothers and O'Donohue 2019). I define hawks as respondents whose military

assertiveness score is in the 75% quantile and doves as respondents whose military assertiveness score is in the 25% quantile.

The results illustrated in Table 5.4 show no evidence for significant treatment heterogeneity. Model 1 includes the binary variable indicating the respondents is a hawk, and Model 2 excludes it. None of the interaction terms between the anger treatment and the four demographic variables have statistically significant coefficient at the 5% level. Turning to foreign policy dispositions, the treatment effect appears heterogeneous to hawkishness only. The interaction between anger and this variable has a negative sign and statistically significant ($p < 0.05$), suggesting that hawks are less susceptible to adversaries' anger expressions. For hawkishness to undermine generalizability, however, average U.S. elites or non-U.S. nationals should be substantively more hawkish than my sample. In conclusion, there is insufficient evidence to believe that the results from the experiment would change substantially in other contexts.

	Model 1	Model 2
Intercept	9.808 (4.964)**	11.117 (4.397)**
Anger	26.581 (6.830)***	23.821 (5.937)***
University education	−5.466 (2.414)**	−3.955 (2.012)**
High income	1.089 (2.503)	−1.071 (2.080)
Age 40-65	−2.218 (2.331)	−2.306 (1.942)
Hawkishness	6.575 (2.575)**	
Isolationism	−1.468 (0.959)	−1.039 (0.815)
White	0.174 (2.734)	2.198 (2.317)
Republican	−4.131 (3.119)	−0.676 (2.566)
Male	−3.102 (2.346)	−1.870 (1.970)
Ideology	2.053 (0.861)**	1.716 (0.741)**
Resolve, T1	0.729 (0.026)***	0.718 (0.023)***
Anger × University education	2.825 (3.388)	2.669 (2.845)
Anger × High income	−3.695 (3.500)	−2.602 (2.952)
Anger × Age 40-65	0.944 (3.385)	−0.849 (2.811)
Anger × Hawkishness	−9.512 (3.768)**	
Anger × Isolationism	1.674 (1.379)	0.434 (1.166)
Anger × White	−3.734 (3.984)	−3.559 (3.304)
Anger × Republican	4.211 (4.512)	−1.952 (3.756)
Anger Anger × Male	−3.209 (3.280)	−4.221 (2.768)
Anger × Ideology	−2.278 (1.240)*	−1.157 (1.057)
R^2	0.598	0.567
Num. obs.	698	1010

***$p < 0.01$; **$p < 0.05$; *$p < 0.1$.

Table 5.4: Heterogeneous treatment effects of the anger expression treatment OLS regression results are shown. The results show that only hawkishness decreases individuals' responsiveness to an adversary's expression of anger. Model 1 includes the hawkishness variable, and Model 2 excludes it because of missing data due the the way I defined hawks and doves.

Chapter 6

Experimental Evidence of the Effect of Anger on Reputational Consequences

The previous chapter has used a survey experiment to study anger as a mechanism through which threats are perceived by their targets: by exogenously manipulating information about the threatening leader's emotions in an experimental vignette, we are able to make causal inferences about the relationship between the credibility of a threat and anger. However, the experiment examines one specific pathway, namely, how anger expressions shape beliefs about casualty sensitivity. In this chapter, I use another survey experiment to offer evidence for the other theorized pathway, in which leader expressions of anger lead to heightened reputational concerns. Yet, the current experiment differs from the previous one in that it examines how members of the public assess the consequences of their leader's behavior in anger on their own country's reputation, rather than their perceptions about an adversary's reputation.

I begin by elaborating on my rationale behind evaluating the hypothesis about anger's reputational mechanism in the domestic politics setting. Reputation at the international stage is a first-order belief, simply a belief foreign observers hold

about a state's or leader's behavioral tendency. As Brutger and Kertzer (2018) suggest, reputation in domestic politics is a second-order belief, which is what citizens think foreign observers think about their country's or leader's behavioral tendency. Studying reputation from the domestic perspective is important because a number of theories rely on this second-order belief about reputations as an implicit or explicit mechanism to explain leader behavior in international crises and conflicts (Fearon 1997; Kurizaki 2007; Smith 1998; Gelpi and Grieco 2015; Guisinger and Smith 2002; Potter and Baum 2010; Mattes and Weeks 2019). Nonetheless, while Brutger and Kertzer (2018) show that the public has a general "taste" for reputation, we understand little about how leaders can manipulate the public's taste to generate strategic advantages in international crises. The results of the current experiment will offer broad implications of anger for the literature on domestic audience cost as well as theories about reputation concerns.

Focusing on domestic politics has direct relevance to testing my theory because reputational consequences of threats emerge not only internationally but also domestically: domestic audiences sanction leaders who tarnish the country's reputation abroad. Importantly, the first- and second-order beliefs need not match. Even if anger does not influence reputations in the eyes of international observers, it could still affect the domestic public's beliefs about the emotion's reputational consequences. Thus, finding evidence with regards to second-order beliefs about reputation will offer support for my theory, regardless of the "true" state of the world internationally. The opposite could also happen, and the public might not understand or care about the impact anger makes on the country's reputations abroad. Given the importance of electoral incentives on leaders' foreign policy behavior, failing to find evidence here might provide a strong counter-evidence for this mechanism. Future research should investigate both first- and second-order beliefs to provide a more comprehensive answer to the question.

Examining domestic audiences introduces important considerations. Since members of a domestic audience have diverse attitudes and preferences (Chaudoin 2014; Kertzer et al. 2014), we might expect significant heterogeneity in citizens' reputational judgments depending on their foreign policy attitudes. For instance, Brutger and Kertzer (2018) argue that hawks and doves view the reputational consequences of the leader's "inconsistnecy" and "belligerence" through a fundamentally different logic. The leader acts inconsistently when she fails to follow through on her threat and belligerently when she initiates a threat in the first place. Hawks, who view the world as threatening and value reputation for credibility as an effective deterrence (Herrmann, Tetlock and Visser 1999), will be highly receptive to leader anger expressions. They will see leader anger as reinforcing reputation when the leader is belligerent and damaging when the leader is inconsistent. Unlike hawks, however, doves understand world politics through a more peaceful lens and do not follow the model of reputation and deterrence. Therefore, doves may see leader anger expressions as neither particularly strengthening nor damaging reputation for honesty when the leader is belligerent or inconsistent. To be consistent with these expectations, I refine Hypothesis 6, discussed in the theory chapter, as follows:

Hypothesis 5a: *An angry leader's inconsistency leads to greater reputational concerns, compared to a calm leader's inconsistency.*

Hypothesis 5b: *An angry leader's belligerence leads to greater reputational benefits, compared to a calm leader's belligerence.*

Hypothesis 5c: *The effect of anger on reputational concerns and benefits is greater for hawks than doves.*

6.1 Experimental Design

6.1.1 Sample

The experiment was conducted in April 2023 on a sample of U.S. public (N = 2,959) recruited via Lucid. The makeup of this experiment's sample appears to resemble that of the previous experiment's. The respondents are 48.2% male, 72.7% Caucasian, and 45.9 years old on average. They are similarly highly educated, with 47.1% received at least a college education. Politically, about 36.6% identified as Republicans, and 43.5% as Democrats. The sample's demographic information is summarized in Table 6.1. As depicted in Figure 6.1, the distributions of foreign policy attitudes and political ideology are also very similar to those of the first experiment.

	Percent Sample
Male	48.20
White	72.60
University education	47.10
Republican	36.60
Democrat	43.50
Independent and other	19.90
Income less than $30,000	34.50
Income $30-60,000	30.30
Income $60-100,000	22.00
Income over $100,000	13.20
Age 18-24	12.50
Age 25-34	18.40
Age 35-44	18.90
Age 45-54	16.10
Age 55+	34.30

Table 6.1: Experiment 2 sample demographics

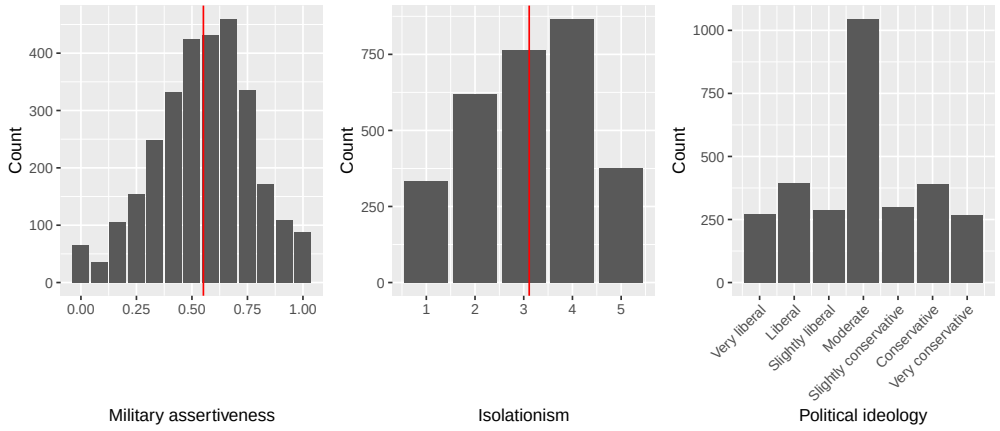

Figure 6.1: Foreign policy attitudes and ideology of the sample

6.1.2 Design

I use the experimental vignette created by Tomz (2007*a*), which now has become the standard in the audience cost literature. I also adopt the extended version from Kertzer and Brutger (2016), who decompose audience costs into "inconsistency cost" and "belligerent cost," and modify it to suit my purposes. The vignette describes a hypothetical extended deterrence scenario, in which the United States president's crisis behavior (stay out, threaten and not engage, and threaten and engage) and emotion (anger and calmness) are varied. The experiment thus uses a 3×2 factorial design, as summarized in Table 6.2. After completing a pre-treatment questionnaire measuring demographic information and foreign policy dispositions, including hawkishness and isolationism, respondents read the following introduction:

> 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.

Following the introduction, respondents were presented with the following

	Anger	No anger
Stay out	1	2
Threaten and not engage	3	4
Threaten and engage	5	6

Table 6.2: Experiment 2 design

background information about the hypothetical vignette:

> A country sent its military to take over a neighboring country in order to get more power and resources. The attacking country's forces killed over 1,000 civilians in the country they are invading. The country being invaded is a democracy and important to U.S. security and economic interests.

The main difference between my version and that of Kertzer and Brutger (2016) is that my version delineates additional, negative information about the attacking country's action during the invasion (killing civilians). The information was provided to ensure that the attacking country would be perceived as the blameworthy party and, more importantly, to prevent the anger treatment from unintentionally influencing beliefs about the characteristics of the attacking country and the attacked country. For example, respondents in the experimental condition where the U.S. president expresses anger at the attacking country might be systematically more likely to believe that the country has committed war crimes. Similarly, if respondents believe that U.S. presidents are more likely to express angry for democracies, which are usually friendly to the U.S. than for nondemocracies, the anger treatment could affect perceptions of the invaded country's regime type and relationship with the U.S. Factoring in explicit information about the attacking country's behavior and the attacked country's characteristics would help address the problem of information leakage (Dafoe, Zhang and Caughey 2018). In addition, like the first experiment in the previous chapter, the president was described as

a male because of gender stereotypes involving gender and anger (Plant et al. 2000).[1]

Respondents were then randomly assigned into one of three conditions that vary the U.S. president's action. Following Kertzer and Brutger (2016), I added the "threaten and engage" condition to the classic two-condition design to distinguish between reputational concerns due to inconsistency and belligerence. In the stay out condition, the U.S. president announces that "U.S. would stay out of the conflict" and fulfills his promise. In the not engage condition, the U.S. president threatens that "if the attack continued, the U.S. military would push out the attackers" but does not follow through when the invading country continues to attack. In the engage condition, the U.S. president delivers the same threat and follows through on it when the attack continues. The engage condition is different from the not engage condition in that the president is inconsistent, holding the threat constant, and differs from the stay out condition in that the president is behaving belligerently by threatening force, holding consistency constant.

Respondents were simultaneously assigned to one of the two conditions manipulating the U.S. president's emotion (anger versus calmness). In the anger condition, respondents were informed that the U.S. president, "who is known for his composure, was furious at the attacking country." Additionally, the president "yelled in anger" when issuing the statement, adding that "he was outraged by the attacking country's actions." By contrast, those in the calmness condition read that the U.S. president, "who is known for his composure, was calm." He also issued the statement "in a measured tone" without any additional comments on his emotions. The president is described as having a composed disposition to highlight the contrast in emotions. The exact wordings for the six treatments are summarized in Table 6.3.

[1] In addition, the fact that the U.S. has never elected a female president means that the male president would be perceived as more realistic.

In all conditions, regardless of whether president sends troops or not, the crisis ends with the attacking country taking over 20-percent of its neighbor's territory. The crisis outcome was held constant to isolate the effects of the president's emotion and behavior. Following Kertzer and Brutger (2016), to further control the crisis outcome, respondents in the engage condition were told that "the U.S. experienced no military casualties." Respondents in the stay out and not engage scenarios did not receive any information about casualties as it is impossible for U.S. troops to experience casualties when they do not fight.

6.1.3 Outcomes

The main outcome variable for this experiment measures the public's concerns about the country's reputation for making honest threats. After the vignette, I measured perceptions of the reputation cost to the U.S. I measured them by asking respondents to report the extent to which they believe "other countries are more or less likely to believe threats made by the U.S" as a result of how the U.S. president handled the situation, on a 7-point scale where 1 means "Extremely less likely," 4 "Neither more or less likely," and 7 "Extremely more likely."[2] With this measure, we can calculate reputational costs due to inconsistency by subtracting reputational concern in the engage condition from reputational concern in the not engage condition. Reputational costs due to belligerence equal reputational concern in the engage condition minus reputational concern in the stay out condition. We can get the total reputational costs by summing the two types of costs.

In addition to measuring reputational concerns, I measured public approval of the U.S. president's handling of the crisis to explore the downstream effect of reputation concerns. If audience cost arises due to the public's concern about reputation (Fearon 1994; Smith 1998), greater reputational costs should translate

[2]For presentation purposes, the variable is rescaled to range from 0 and 1.

Condition	Treatment
Stay out, calm	• The U.S. President, who is known for his composure, was calm. • The President said in a measured tone that the U.S. would stay out of the conflict. • The attacking country continued to invade. The U.S. President did not send troops.
Not engage, calm	• The U.S. President, who is known for his composure, was calm. • The President said in a measured tone that if the attack continued, the U.S. military would push out the attackers. • The attacking country continued to invade. The U.S. President did not send troops.
Engage, calm	• The U.S. President, who is known for his composure, was calm. • The President said in a measured tone that if the attack continued, the U.S. military would push out the attackers. • The attacking country continued to invade. The U.S. President sent troops to stop the invasion.
Stay out, anger	• The U.S. President, who is known for his composure, was furious at the attacking country. • The President yelled in anger that although the U.S. would stay out of the conflict, he was outraged by the attacking country's actions. • The attacking country continued to invade. The U.S. President did not send troops.
Not engage, anger	• The U.S. President, who is known for his composure, was furious at the attacking country. • The President yelled in anger that if the attack continued, the U.S. military would push out the attackers, and that he was outraged by the attacking country's actions. • The attacking country continued to invade. The U.S. President did not send troops.
Engage, anger	• The U.S. President, who is known for his composure, was furious at the attacking country. • The President yelled in anger that if the attack continued, the U.S. military would push out the attackers, and that he was outraged by the attacking country's actions. • The attacking country continued to invade. The U.S. President sent troops to stop the invasion.

Table 6.3: Experiment 2 vignette and treatments

to higher audience costs. I asked respondents to evaluate the president a 7-point likert scale (Strongly disapprove, Disapprove, Somewhat disapprove, Neither approve nor disapprove, Somewhat Approve, Approve, Strongly approve), and used this measure to build a binary variable approval, which takes a value of 1 if the respondent at least somewhat approves and 0 otherwise.

Lastly, to validate that the anger treatment does not work through other mechanisms, I included placebo test questions at the end of the survey. As alluded earlier, respondents might believe that the president is angry because the attacking country is violent or the attacked country is important to the U.S. I thus asked respondents to report whether or not they think "the attacking country committed major war crimes during the conflict," and to what degree they believe "the attacking country's takeover of its neighbor would do to the security and economy of the U.S" (No damage at all, A little bit, A moderate amount, A great amount). It could also be that U.S. citizens believe that the president is more likely to get angry at dictatorships than democracies. To check for this possibility, I asked respondents to indicate whether they believe "the attacking country is a dictatorship or a democracy."

6.2 Results

6.2.1 Manipulation checks and placebo test

The survey includes two manipulation check questions. First, following the scenario, I asked respondents, "to what extent did the U.S. President express anger in his remarks?" and measured their perceptions on a 7-point scale (1: Not at all, 4: Moderately, 7: Extremely). The result of a t-test indicate a successful manipulation of the president's emotion. Respondents in the anger condition ($M = 5.44$, $SD = 1.50$) viewed the president as significantly angrier ($t(2904.2) = 52.2$, $p < 0.001$) than

those in the calmness condition (M = 2.36, SD = 1.70). The second manipulation check assessed whether respondents were able to recall the president's action. About 93.8% of respondents in the engage condition and 91.2% in the not engage and stay out conditions successfully recalled the president's behavior.

Placebo tests provide some evidence that the anger treatment does not alter other beliefs related the crisis. The first two placebo questions show that the vignette successfully held constant the attacking country's violence and the attacked country's relationship with the U.S. On average, respondents in the anger condition were no more or less likely to believe that the attacking country committed a war crime ($t(2954.7) = 0.66$, $p < 0.50$) and that the attacked country is important to U.S. interests ($t(2959.9) = 0.75$, $p < 0.46$). However, I find that more respondents who receive the anger treatment believe that the attacking country is a dictatorship rather than a democracy ($t(2935.4) = 2.01$, $p < 0.04$).

6.2.2 Anger increases reputational concerns

Emotion	Inconsistency	Belligerence	Total
Anger	-0.428	0.339	-0.128
Calm	-0.381	0.259	-0.122
Difference	-0.047***	0.08***	-0.006

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$.

Table 6.4: Effect of anger on reputational concerns due to inconsistency and belligerence (full sample)

I first present the joint treatment effects of anger and inconsistency and belligerence on the respondents' reputational concerns on the whole sample. The results are summarized in Table 6.4. The quantities of interest are the difference in the measure of the country's credibility calculated from 3,000 bootstraps. Negative values represent reputational costs and positive values mean reputational benefits. The results provide strong support for my expectations that anger expressions

increase the reputational stake of making threats. When the president expresses anger, inconsistency results in the perceived cost to the reputation of 0.428. When the president appears calm, the cost is 0.381, representing a difference of 0.047 ($p < 0.05$). In sum, the perceived reputational cost is about 11% higher when an angry president acts inconsistency by failing to carry out his threat relative to when a calm president acts inconsistently. This confirms Hypothesis 6a.

I find opposite, but similarly large effects for the reputational concern for a belligerent president. Examining the third column in Table 6.4, both anger and calm conditions have positive values, meaning that the respondents perceived reputational benefits in belligerence in anger. In the anger condition, the perceived benefit to the country's reputation is 0.339, and in the calm condition, the benefit is 0.259. This represents about a 24% increase in reputational benefits ($p < 0.01$). This finding confirms Hypothesis 6b and is consistent with my contention that leaders can develop a reputation for being easily angered by fulfilling threats while expressing anger, which provides credibility to future threats.

I did not make a prediction about the aggregated reputational cost (stay out minus threaten and not engage, or the sum of inconsistency and belligerence costs) because I expected that the reputational consequences of inconsistency in anger and belligerence in anger would have countervailing effects. The null results shown in the fourth column of Table 6.4 confirms this expectation. Whether the president is perceived as anger or calm does not appear to have a statistically significant effect on the total reputational concern is ($p = 0.378$). It should be highlighted that significant heterogeneity due to inconsistency and belligerence exists behind the null effect. Focusing on the aggregate reputational effect would lose sight of the heightened reputational concern that arises when an angry president is inconsistent and the perceived reputational benefits that stem from an angry president's belligerence.

6.2.3 Do hawks and doves respond to leader anger differently?

Do foreign policy hawks and doves respond differently to angry behavior from their leader? To investigate this question, I split the sample into hawks and doves, with hawks defined as respondents whose military assertiveness scores are in the top quartile and doves defined as respondents whose scores are in the bottom quartile. I then re-calculated the reputational concerns.

The results, displayed in Tables 6.5 and 6.6, indicate support for Hypothesis 6c. Hawks perceive the reputational cost of 0.513 as a result of the president's failure to follow through on his threat when the president behaves in anger, and perceive the cost of 0.41 when the president is calm, resulting in a 20% increase in reputational cost ($p < 0.01$). Hawks also believe that belligerence in anger improves the country's reputation. The effect of the anger treatment on the perceived reputational benefits from belligerence is 0.169 ($p < 0.01$). This represents a 36.4% increase in the beliefs that belligerence improves the country's reputation, an effect that is substantially large and highly statistically significant.

On the other hand, doves appear to be less receptive to the president's anger. As shown in Table 6.6, doves perceive neither greater reputational costs due to inconsistency nor greater reputational benefits due to belligerence when the president expresses anger. These results are consistent with evidence that hawks are systematically more concerned about the country's reputation for credibility than doves (Brutger and Kertzer 2018). The fact that hawks responded to anger but doves did not provides suggestive evidence that anger expressions alter reputational consequences of a threat.

Emotion	Inconsistency	Belligerence	Total
Anger	-0.513	0.464	-0.134
Calm	-0.41	0.295	-0.115
Difference	-0.103***	0.169***	-0.019

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$.

Table 6.5: Effect of anger on reputational concerns due to inconsistency and belligerence (hawks only)

Emotion	Inconsistency	Belligerence	Total
Anger	-0.397	0.267	-0.121
Calm	-0.363	0.256	-0.107
Difference	-0.034	0.011	-0.014

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$.

Table 6.6: Effect of anger on reputational concerns due to inconsistency and belligerence (doves only)

6.2.4 Does anger increase audience costs?

Lastly, I explore whether the public's increased reputational concerns due to leader anger ultimately lead to higher audience costs. Although definitively answering this question is beyond the scope of the current study without a proper mediation analysis design, investigating this link will be useful because it will suggest another mechanism through which anger expressions work to increase the credibility of threats. Building upon the finding in the previous section that hawks and doves have different reactions to the leader's anger, I conduct the same sub-sample analysis using hawks and doves to calculate inconsistency costs, belligerence costs, and audience costs.

The analysis provides mixed results for the effect of leader expressions of anger on domestic political disapproval. The results for the hawks sample are shown in Table 6.7. For hawks, the president's anger does not seem to affect the total audience costs, calculated as the difference in audience costs in the stay out condition and the not engage condition. However, the anger treatment increases inconsistency costs by 15 percentage points ($p = 0.055$) and increases belligerence

Emotion	Inconsistency cost	Belligerence cost	Audience cost
Anger	-0.612	0.512	-0.137
Calm	-0.519	0.414	-0.105
Difference	-0.093*	0.098*	-0.032

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$.

Table 6.7: Effect of anger on audience cost and its components (hawks only)

Emotion	Inconsistency cost	Belligerence cost	Audience cost
Anger	-0.265	0.073	-0.171
Calm	-0.348	0.227	-0.121
Difference	0.083	-0.154**	-0.05

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$.

Table 6.8: Effect of anger on audience cost and its components (doves only)

benefits by 19 percentage points ($p = 0.068$), although the differences are only statistically significant at the 10% level. Similarly, for doves, the anger treatment has no significant impact on audience costs and inconsistency costs they impose on the leader. However, I find that doves react negatively to the president's anger. The mean approval for a president acting belligerently in the anger condition is 7.3%, and is 22.7% in the calm condition, and this effect achieves statistical significance at the conventional level ($p < 0.05$). The overall pattern of findings is consistent with my expectations about how anger affects the way hawks and doves would perceive reputational consequences of inconsistency and belligerence, hinting that the effect of leader anger on reputational concern has downstream effects on audience costs. However, given the weak statistical significance, I conclude that there is only weak support for the idea that anger expressions shape audience cost or its components.

Chapter 7

Conclusion

This dissertation began by asking why some threats are more credible than others. There is a widespread agreement among scholars and decision makers that threats are an important instrument of coercion that helps to obtain favorable outcomes in foreign policy crises and disputes without escalating into war. Yet, although scholars of crisis diplomacy have put forth a number of explanations for what makes threats credible, the literature as a whole has yet to resolve the controversies surrounding existing explanations. Moreover, in studying threat effectiveness, scholars largely ignored diplomatic statements that contain the threats, missing a significant source of variation that could explain credibility. This has left decision makers who seek to achieve peaceful coercion with incomplete understandings about what could allow them to make credible threats and how to interpret warnings from foreign adversaries.

I also explained why the traditional focus on costly signaling might only tell half of the story. Because most theories of costly signaling are developed from the sender's perspective, they rarely problematize the assumption that diplomatic messages are perceived and interpreted in a perfectly rational, objective fashion. Moreover, signaling theories tend to focus on contextual variables, such as do-

mestic politics and audience costs, and require additional assumptions about the target's knowledge about the sender's characteristics that are often not directly observable. Even if we accept these assumptions, many prominent contextual indicators like regime type are relatively unchanging. This means that they leave many interesting cases unaccounted for, such as repeated crises, threats between states with similar political institutions, and different success rates from the same leader's threats.

To address these shortcomings, I argued that we can leverage variation in the statements leaders make to explain the credibility and effectiveness of threats. In doing so, I put forth leader emotionality as an under-appreciated yet critical determinant of credibility. Drawing on social psychology as well as evolutionary approaches to emotions, I explained how leaders' anger conveyed in their threatening statements greatly influence how targets evaluate the threats contained in the statements. Expressions of emotions communicate a person's dispositions, goals, beliefs, and preferences, and observers use the emotions to make inferences about the person's future intentions. What information anger communicates? I proposed two specific pathways. First, anger decreases the perceptions of the leader's sensitivity to the cost of conflict. Evolutionary psychologists view anger as a result of human adaptation to solve bargaining problems (Sell 2011). Anger is an emotion that orchestrates various cognitive and bodily mechanisms in the angry person to commit them to aggression in order to incentivize the target to improve how it treats the person. This strategic value of anger plays an important role in coercive diplomacy.

The second pathway is that expressions of anger increase the reputational consequences of threats. One reason why anger expressions affect reputational consequences is that leaders who are seen as angry raise expectations about their commitment to carrying out their threats. Failing to follow through an angry threat

will betray others' expectations to a greater extent, resulting in a larger loss in the reputation for honest threats. By successfully fulfilling threats delivered in anger, on the other hand, leaders set and meet high expectations about their commitments to using force, gaining stronger reputations. Another reason why angry leaders can put reputations at stake to a greater extent than leaders who are not seen as angry is that leaders may be able to develop reputations for anger. Reputations are multidimensional and can arise regarding any behavioral trait or disposition (Dafoe, Renshon and Huth 2014). Reputations for being provoked easily or for becoming furious over a particular issue can provide strategic advantages in crisis diplomacy. This mechanism is different from the first one in that it requires an assumption, albeit weaker, about the target's ability to reason backward about the angry leader's strategic motivations.

The fundamental intuition behind my argument is that emotion expressions are an intuitive indicator of a leader's resolve. Unlike theories of costly signals, my theory about leader anger does not rest on the strong assumption that leaders are perfect information processors. I argued that the logic of anger is understood at both the sending and the receiving ends of the signaling mechanism, and this claim is grounded on empirical evidence that humans likely share a common understanding of the socio-functional meaning of anger. I also explained how the automatic nature of emotions leads them to be relied upon as an indicator of intentions. While emotions, like any other indicators, are not totally free of misinterpretation and manipulation, they carry significant inherent credibility because their genuine expressions are believed to be beyond full conscious control. To be sure, this is not to suggest that other strategic and contextual factors do not matter. My contention is that, as Katagiri and Min (2019) put it, "elites are literally incapable of processing and acting upon every piece of information running across their desks" and thus will latch onto the most readily interpretable source of

information before assessing more complicated ones.

The empirical chapters of this dissertation tested my arguments through an analysis of world leaders' public statements as well as two survey experiments. Chapter 3 described in detail how I collected and coded the data on public statements. I collected historical records of radio or television broadcasts of world leaders' public statements between 1946 and 1996. This original dataset represents the first quantitative measure of elite emotional expressions across and and space, a significant methodological advance for the scholarship on coercive diplomacy and the study of emotions in world politics. The data allows scholars to use existing data on international crises to understand how elite statements of anger and threats affect the way crises unfold, in a temporal and spatial scope that was not feasible with case studies or experiments.

7.1 Summary of Findings

Across three studies, I found powerful support for my theory. In Chapter 4, I used the data on public threats and anger statements to test the hypothesis that threats embedded in anger expressions are more effective in the context of international crises. I used a statistical analysis to model the outcome of a crisis as a function of the interaction between the level of threat and the level of anger expression contained in the leaders' public statements. The results supported my expectation, even after controlling for alternative explanations such as regime type, relative power, and the level of violence used in the crisis. The analysis paid a meticulous attention to ensuring that the independent variable captures anger specifically and not other emotions. I included covariates accounting for the overall emotionality of the statement, and I conducted a series of supplementary analyses to demonstrate the robustness of my measure of anger expressions and the resulting findings. One

of the major findings of the analysis is that issuing threats makes a meaningful impact on the probability of winning only when the threats are accompanied by relatively high levels of anger. Therefore, anger does not simply improve credibility; rather, it might be an essential mechanism for threats to be perceived as credible. I pursued this argument further in the next chapter.

Chapter 5 used a survey experiment to evaluate my theory. Despite my careful efforts to rule out alternative explanations in the previous chapter, the use of the observational data did not allow me to eliminate the issue of confounding. The experiment in this chapter, conducted on a sample of U.S. public, randomized assignment into expressions of anger to isolate the effect of anger on beliefs about resolve. Using a hypothetical crisis scenario between the U.S. and Russia, the results strongly confirmed my hypotheses. One major finding is that the Russian leaders' anger expressions significantly increased the respondents' beliefs about the leader's resolve. Examining how respondents updated their beliefs about resolve also confirmed the hypothesis that threats are seen are more credible when the threatening leader is angry than when the same leader is not angry. I found that public threats from the Russian leader had negative effects on beliefs about Russia's resolve. This is striking evidence speaking to the crucial role leader anger plays as a mediator of threats.

Chapter 6 used another survey experiment to test my argument about reputations. In doing so, I focused on the domestic audiences' perceptions about reputational consequences. Confirming Hypothesis 6, I found significant effects of leader anger expressions. U.S. respondents reported greater concerns about U.S. reputations for honest threats in response to an angry president's inconsistency, compared to a calm president's inconsistency. The respondents also indicated greater perceptions of reputational gains in response to the president's belligerence when the president was angry than when the president was calm. Exploratory

analyses showed that this effects of anger are greater for foreign policy hawks and doves, providing additional support for the idea that leader anger works through a mechanism related to reputations. Yet, I only found weak support for the subsequent investigation of whether anger expressions raise audience costs for leaders. Hawks, but not doves, were more disapproving of the president for inconsistency and more approving of the president for belligerence when the president expressed anger, and these effects were statistically significant only at the 10% level.

7.2 Implications

The findings from this study provide important implications for scholars of coercive diplomacy and foreign policy practitioners. On the most general level, I have shown throughout the dissertation that we need to take leaders' statements more seriously in order to fully understand why some threats succeed while others fail. Statements have received far less attention than they deserve in the current literature, which has primarily focused on the contextual factors that shape the costs associated with issuing or backing down from threats. To my best knowledge, only Trager (2017) and McManus (2017) explicitly theorize about variation within statements that can explain the credibility of threats. To be sure, contextual factors are important since they help us make broad predictions about under which conditions we would expect threats to be credible. However, as this dissertation shows, understanding what leaders say in their statements and how they say it are equally, if not more, important in explaining credibility, which allows us to leverage greater variation at the elite level instead of relying on macro-level proxies such as regime type. A more comprehensive understanding of threat credibility requires theorizing about statements as well as how the statements

interact with contextual factors.

More specifically, the findings on emotion expressions challenge the way we traditionally study coercive bargaining and credible signaling. The current literature has generated a great deal of insights into conditions for costly, and thus, credible threats, but the relative importance of these factors has not been systematically investigated. We particularly understand little about how the difficulty of inference can impact a signal's perception. Along the same lines, while research on costly signaling and misperception highlights distinct hurdles faced by senders and receivers in communicating resolve, few studies have incorporated insights from the two approaches. This dissertation took up this task to add to a more comprehensive understanding of threat credibility. My arguments about emotional expressions explicitly take into consideration how the intuitiveness of a signal affects its perception and credibility, and my empirical results show that threats are largely incredible without anger even when they are issued publicly.

I am not the first to suggest that emotions and affect facilitate signaling. As introduced in the introductory chapters, studies on interpersonal diplomacy investigate how leaders rely on their personal impressions of others as an intuitive way to assess their intentions rather than costly signals (Yarhi-Milo 2013; Holmes 2013; Hall and Yarhi-Milo 2012; Wong 2016). My work pursues this line of inquiry but innovates by pinpointing what about personal impressions—anger expressions—conveys resolve and providing both causal and generalizable evidence for it.

This dissertation's framework broadens our understanding of how emotions can influence coercive diplomacy. Existing work on anger in international relations tends to portray anger as a cause of coercion failure and subsequent conflict (e.g. Powers and Altman 2023; Hall 2017; Dafoe, Hatz and Zhang 2021). However, existing studies focused on *experienced* anger. By examining the effects of *expressed*

anger, I found that anger can prevent conflict by persuading adversaries to retreat.[1] The findings of this dissertation warrant further research to explore the full effects of emotional expressions as well as their interplay with experienced emotions in shaping the dynamics of crisis bargaining.

On the methodological front, this dissertation offers new data on leader's public statements over time and space. Note that I label each statement's threats and anger expressions to test my theory, but the text data contain the raw verbatim statements about crisis adversaries. The data could be labeled using other methods, be it hand-coding, machine learning models, or automated content-analysis, to detect and analyze other interesting variation in leaders' statements. For example, scholars might be able to use the data to explore how different parts of the speech interact with each other or how leaders address the problem of multiple audiences when issuing public statements. Scholars could also use the data to develop quantitative measures of leaders' personal dispositions indirectly revealed in their speeches. Therefore, the statements data will make contribution to the study of coercive diplomacy as well as the broader literature on the importance of individual leaders in world politics.

This dissertation also provides practical insights for foreign policy. The results can inform policy makers who seek to deter or compel an adversary through threats to use force. Empirical findings from three empirical studies provide converging evidence that threats are credible when the threatening leaders are viewed as sufficiently angry, and that threats could even be perceived as a sign of lack of resolve when threats are devoid of anger. The conclusions from these studies thus suggest the importance of choosing the appropriate tone and rhetoric when issuing threats. For example, leaders might try to suppress their genuine emotions and display calmness in crises in hopes of reassuring the public, but the

[1] For similar approaches, see Hall (2015) and Wong (2019).

trade-off is that the calm threats might not be believed by adversaries. Of course, I do not suggest that leaders should act on their emotions all the time; rather, the conclusion should be that emotional honesty can be strategically prudent.

The same lessons can be drawn for policy makers who wish to assess threats from other countries. Carefully analyzing the emotionality of an adversary's rhetoric would help evaluate the credibility of its threats and warnings. In addition, this dissertation used public statements translated into English by the CIA personnel and discussed the qualifications of the analysts and translators involved in the original data collection efforts. For the intelligence community, the work presented here highlights the importance of recruiting and training experts in foreign countries who are able to understand the nuances of emotional expressions. In addition, the findings also speak to the importance of diplomats' and decision makers' ability to deal with possible attempts for emotional deception.

7.3 Avenues for Future Research

This dissertation project opens up several promising avenues for future research. Exploratory analyses have provided suggestive evidence that anger expressions increase perceptions of domestic audience costs of making empty threats. This suggests a theoretically important mechanism in which the size of audience costs varies based on whether a leader displayed anger or not. While I only found weak empirical support for this possibility, this question is worth investigating further given the centrality of audience costs in the field of International Relations.

Another potential direction is studying the role of the media. While this study used an online archive of transcribed speeches, in modern times, foreign observers may be more likely to learn about a leader's statements indirectly through media reports than through listening to the statements directly. In addition, given that

leaders increasingly use social media to communicate their intentions (Harris and Lin-Greenberg 2024), a follow-up study could examine the extent to which the medium of emotional communication moderates the effects of emotion expressions. Does the use of social media help leaders communicate their emotions more easily to wider audiences? Are targets more likely doubt the sincerity of emotions conveyed through social media since the physiological elements of emotional expressions cannot be observed?

In a similar vein, it would be of great interest to examine how dispositional and contextual factors alter the effects of anger expressions. One important factor is gender. The survey experiments presented in Chapters 5 and 6 held the leader's gender constant at male. However, as the number of female leaders is rapidly increasing around the world, it would be important to explore how gender stereotypes might influence the effects of anger expressions. It could be that targets take female leaders' anger less seriously because women expressing anger tend to be perceived as incompetent due to the gender stereotype. Conversely, anger from female leaders might be particularly informative precisely because of these social costs associated with it. A third possibility might be that the stereotype that women are easily swayed by emotions amplifies the effect of anger on perceptions of cost sensitivity and beliefs about resolve. These are empirical questions worth investigating using experimental data. In addition, by collecting more observational data on statements, it would be interesting to study how male and female leaders differ in their conduct of "emotional diplomacy" (Hall and Ross 2015), not just regarding anger but also other emotions such as sadness.

An interesting candidate for a contextual moderator is regime type. Elected democratic leaders may be more constrained in their foreign policy than authoritarian rulers because of stronger institutional checks and electoral accountability (Croco and Weeks 2016). Although democratic leaders have considerable power in

security matters, decisions are typically made in a group setting like the National Security Council. However. nondemocratic leaders, and especially personalist and military dictators (Weeks 2014), might wield considerably greater individual power over decisions to use force. Therefore, nondemocratic rulers might be viewed as more susceptible to the influence of the emotions. This suggests that the effects of anger expressions on threat credibility might generally be larger for nondemocratic rulers than democratic leaders. This claim is contrary to the conventional wisdom that democracies make more credible threats than other regime types, and might possibly explain why recent experiments failed to find significant evidence for a democratic advantage in threat effectiveness (Kertzer, Renshon and Yarhi-Milo 2021; Yarhi-Milo, Kertzer and Renshon 2018).

Another contextual variable worth examining is power. Anger expressions from relatively more powerful leaders might be more effective. Evidence from bargaining experiments suggests that anger tends to evoke fear in adversaries when expressed by high-power individuals but causes counter-anger when expressed by low-power individuals (Lelieveld et al. 2012). In addition, higher-power bargainers may be less motivated to pay attention to their opponents' emotions (Van Kleef et al. 2006). In crisis bargaining, anger expressions should be more likely to evoke fear in the adversary when the leader is militarily more powerful. The adversary's decision-makers and the public might become fearful of war and inclined to back down. On the other hand, anger from less powerful leaders could provoke adversaries by making it difficult for them to "save face" while backing down. Future study should investigate the extent to which militarily superior leaders are better able to utilize anger expressions to make more credible threats credible.

Future work should also study whether other emotions have meaningful consequences for coercive diplomacy. The current project focuses on anger, but leaders reveal in their statements a variety of other emotions, including fear,

sadness, and surprise. And this allows us to ask questions about broader impacts of emotions. For example, does conveying fear of a nuclear war, intentionally or not, help convince the adversary of the risk of escalation, or does it signal lack of resolve? Does expressing surprise in response to a challenge inform the adversary that its challenge was successful, emboldening it to make greater demands? A systematic theorization and an empirical investigation are necessary to shed light on these questions.

Furthermore, the work here highlights the utility of leaders' statements as text data. Scholars of private diplomacy suggest that effective emotional communications require interpersonal interactions (Holmes 2013; Wong 2016). Yet, I used original data on public statements to show that emotional expressions have meaningful effects even in public diplomacy. Future work could collect a wider variety of statements, including private diplomatic communications, to understand how the patterns and the effects of anger expressions vary depending on the channel of communications. Doing so will broaden our understanding of the conditions under which emotional expressions are effective, and also contribute to the literature on the relative effectiveness of public and private threats.

Finally, although the empirical focus on this project is on crisis diplomacy, anger might be relevant in other non-crisis bargaining situations that involve non-human costs. For example, anger expressions during an international trade dispute could signal increased willingness to bear economic costs for more a favorable negotiated settlement. Future research should investigate such possibilities to examine the scope of the argument offered here.

Chapter 8

Appendix

8.1 Appendix 1: Survey Instrument for Experiment 1

Pre-treatment covariates

- [Foreign policy attitudes] Please tell us how much you agree with each of the following statements. (Strongly disagree, Disagree, Somewhat disagree, Neither agree nor disagree, Somewhat agree, Agree, Strongly agree)
  - [Militant Internationalism 1] The best way to ensure world peace is through American military strength
  - [Militant Internationalism 2] The use of military force only makes problems worse (reverse coded)
  - [Militant Internationalism 3] Going to war is unfortunate, but sometimes the only solution to international problems
  - [Isolationism] The U.S. needs to play an active role in solving conflicts around the world
  - [Attention check 1] Please click ”Neither agree nor disagree”
  - [Attention check 2] World War I came after World War II

- [Ideology] In general, do you think of yourself as ... (Very liberal, Liberal, Slightly liberal, Moderate, Slightly conservative, Conservative, Very conservative)

Baseline vignette

- This survey is about world politics. We are going to describe a situation our country has faced many times in the past and will probably face again, and then ask your opinions about it. Do you agree to read the details carefully and give thoughtful answers? (No, Yes)[1]

- **Here is the situation:** In 2040, the United States is involved in a dispute with Russia. Russia is not a democracy and has a strong military. The dispute began with a collision between a U.S. shipping vessel and a ship registered to Russia. During the collision, injuries were reported on both sides. Additionally, both countries maintain that their ship was carrying sensitive military technology, and are suspicious of the motives of the other side, leading to a tense standoff at sea. Both countries' leaders have not yet made any statements regarding the dispute.

Outcome Variables

- [Probability] Countries may use force to resolve disputes like these. Given the information available, what is your best estimate about whether Russia will use force in this dispute, ranging from 0% to 100%? (Text entry)

- [Uncertainty] How certain are you of the estimate you gave above, on a scale from 0% to 100%? (Text entry)

[1] Answering "No" to this question fails the third attention check.

- [Casualties] Given the information available, what is your best estimate about how many military casualties Russia's leader is willing to accept to win this dispute? (Russia's leader will not accept any military casualties./3 military casualties/30 military casualties/300 military casualties/3,000 military casualties/30,000 military casualties/Russia's leader will accept any number of casualties.)

- [Other mediators] Given the information available in this scenario, please tell us how much you agree with each of the following statements about Russia. (Strongly disagree, Disagree, Neither agree nor disagree, Agree, Strongly agree)

- [Honor] Russia's leader intends to maintain Russia's national honor.

- [Stake] The military technology the Russian ship was carrying is valuable for Russia.

- [Disapproval] Russia's leader will face strong public disapproval if Russia does not use force in this dispute.

Treatments

- **Here is how the Russian leader responds to the dispute:**

- [Threat, Anger] Russia's leader explodes in rage after hearing about the dispute at sea. In a public statement, he yells in outrage that they will use force if the U.S. does not back down and that he will "make it clear that we are angered by how America treated us."

- [Threat, No anger] Russia's leader does not seem angry at all after hearing about the dispute at sea. In a public statement, he says in a calm, measured voice that they will use force if the U.S. does not back down.

Placebo test questions

- [Emotions] Please rate how much each of the following words describes how you feel when thinking about this scenario. (Very slightly or not at all, A little, Moderately, Quite a bit, Extremely)
  - [Fear] afraid, scared, frightened, nervous, jittery, shaky
  - [Hostility] angry, hostile, irritable, scornful, disgusted, loathing
- [Power] Russia has a strong military. (Strongly disagree, Disagree, Neither agree nor disagree, Agree, Strongly agree)

Manipulation checks

- [Threat manipulation] In the scenario we just described, what does the Russian leader end up doing? (Russia's leader signs a trade agreement with the U.S., Russia's leader threatens to use force if the U.S. does not back down.)
- [Anger manipulation] In the scenario we just described, on a scale of 1-5, how angry do you think the Russian leader is? (1: Not at all angry, 3: Moderately angry, 5: Extremely angry)

Post-treatment attention check

This question appears in the same matrix as "Other mediators."

- [Attention check] Please click "Disagree"

8.2 Appendix 2: Survey Instrument for Experiment 2

Pre-treatment covariates

- [Foreign policy attitudes] Please tell us how much you agree with each of the following statements. (Strongly disagree, Disagree, Somewhat disagree, Neither agree nor disagree, Somewhat agree, Agree, Strongly agree)
  - [Militant Internationalism 1] The best way to ensure world peace is through American military strength
  - [Militant Internationalism 2] The use of military force only makes problems worse (reverse coded)
  - [Militant Internationalism 3] Going to war is unfortunate, but sometimes the only solution to international problems
  - [Isolationism] The U.S. needs to play an active role in solving conflicts around the world
  - [Attention check 1] Please click "Neither agree nor disagree"
  - [Attention check 2] World War I came after World War II
- [Ideology] In general, do you think of yourself as ... (Very liberal, Liberal, Slightly liberal, Moderate, Slightly conservative, Conservative, Very conservative)

Vignette

[Introduction]

- 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 your opinion about it.

[Slide 1]

- A country sent its military to take over a neighboring country in order to get more power and resources.
- The attacking country is led by a [democratically elected leader/dictator].
- The country being invaded is friendly to the U.S. but not a U.S. ally.

[Stay out, Anger]

- The [Republican/Democratic] U.S. President became furious when he was told of the attack.
- "Although the U.S. will stay out of the conflict," he said angrily in a televised statement, "I am outraged by the attacking country's aggression against its neighbor!"
- The attacking country continued to invade. In the end, the President did not send troops, and the attacking country took over 20 percent of its neighbor's territory.

[Threaten force and not engage, Anger]

- The [Republican/Democratic] U.S. President became furious when he was told of the attack.
- "I am outraged by the attacking country's aggression against its neighbor," he said angrily in a televised statement. "If the attack continues, the U.S. military will push out the attackers!"

- The attacking country continued to invade. In the end, the President did not send troops, and the attacking country took over 20 percent of its neighbor's territory.

[Threaten force and engage, Anger]

- The [Republican/Democratic] U.S. President became furious when he was told of the attack.
- "I am outraged by the attacking country's aggression against its neighbor," he said angrily in a televised statement. "If the attack continues, the U.S. military will push out the attackers!"
- The attacking country continued to invade. In the end, the President sent troops to stop the invasion, and the attacking country took over 20 percent of its neighbor's territory.
- The U.S. experienced no military casualties.

[Stay out, No anger]

- The [Republican/Democratic] U.S. President seemed to have no emotional reaction when he was told of the attack.
- "The U.S. will stay out of the conflict," he said calmly in a televised statement.
- The attacking country continued to invade. In the end, the President did not send troops, and the attacking country took over 20 percent of its neighbor's territory.

[Threaten force and not engage, No anger]

- The [Republican/Democratic] U.S. President seemed to have no emotional reaction when he was told of the attack.

- "The U.S. military will push out the attackers if the attack continues," he said calmly in a televised statement.
- The attacking country continued to invade. In the end, the President did not send troops, and the attacking country took over 20 percent of its neighbor's territory.

[Threaten force and engage, No anger]

- The [Republican/Democratic] U.S. President seemed to have no emotional reaction when he was told of the attack.
- "The U.S. military will push out the attackers if the attack continues," he said calmly in a televised statement.
- The attacking country continued to invade. In the end, the President sent troops to stop the invasion, and the attacking country took over 20 percent of its neighbor's territory.
- The U.S. experienced no military casualties.

Outcome Variables

- [Reputation] As a result of the President's handling of the situation, on a scale of 1-7, do you think other countries are more or less likely to believe threats made by the U.S.? (1: Extremely less likely, 4: Neither more nor less likely, 7: Extremely more likely)
- [Approval] How much do you approve or disapprove of the way the President handled the situation? (Strongly disapprove, Disapprove, Somewhat disapprove, Neither approve nor disapprove, Somewhat approve, Approve, Strongly approve)

Manipulation check

- [Anger manipulation check] Based on the President's remarks, how angry do you think he was? (1: Not angry at all, 3: Moderately angry, 5: Extremely angry)

Bibliography

Acharya, Avidit, Matthew Blackwell and Maya Sen. 2018. "Analyzing Causal Mechanisms in Survey Experiments." *Political Analysis* 26(4):357–378.

Alonso-Arbiol, Itziar, Fons JR van de Vijver, Itziar Fernandez, Dario Paez, Miryam Campos and Pilar Carrera. 2011. "Implicit Theories about Interrelations of Anger Components in 25 Countries." *Emotion* 11(1):1.

Art, Robert J and Kelly M Greenhill. 2018. "The Power and Limits of Compellence: A Research Note." *Political Science Quarterly* 133(1):77–97.

Averill, James R. 1982. *Anger and Aggression: An Essay on Emotion*. Springer Science & Business Media.

Aviezer, Hillel, Shlomo Bentin, Veronica Dudarev and Ran R Hassin. 2011. "The Automaticity of Emotional Face-Context Integration." *Emotion* 11(6):1406.

Bassan-Nygate, Lotem, Jonathan Renshon, Jessica LP Weeks and Chagai M Weiss. 2023. "The Generalizability of IR Experiments Beyond the US." *APSA Preprints* .

Berinsky, Adam J, Michele F Margolis, Michael W Sances and Christopher Warshaw. 2021. "Using Screeners to Measure Respondent Attention on Self-Administered Surveys: Which Items and How Many?" *Political Science Research and Methods* 9(2):430–437.

Berkowitz, Leonard. 1990. "On the formation and regulation of anger and aggression: A cognitive-neoassociationistic analysis." *American psychologist* 45(4):494.

Berkowitz, Leonard and Eddie Harmon-Jones. 2004. "Toward an Understanding of the Determinants of Anger." *Emotion* 4(2):107.

Bleiker, Roland and Emma Hutchison. 2008. "Fear No More: Emotions and World Politics." *Review of international studies* 34(S1):115–135.

Brecher, Michael and Jonathan Wilkenfeld. 1997. *A Study of Crisis*. University of Michigan Press.

Brutger, Ryan and Joshua D Kertzer. 2018. "A dispositional theory of reputation costs." *International Organization* 72(3):693–724.

Brutger, Ryan, Joshua D Kertzer, Jonathan Renshon, Dustin Tingley and Chagai M Weiss. 2023. "Abstraction and Detail in Experimental Design." *American Journal of Political Science* 67(4):979–995.

Bullock, John G and Donald P Green. 2021. "The Failings of Conventional Mediation Analysis and a Design-Based Alternative." *Advances in Methods and Practices in Psychological Science* 4(4):25152459211047227.

Burgoon, Judee K. 1993. "Interpersonal Expectations, Expectancy Violations, and Emotional Communication." *Journal of Language and Social Psychology* 12(1-2):30–48.

Busso, Carlos, Zhigang Deng, Serdar Yildirim, Murtaza Bulut, Chul Min Lee, Abe Kazemzadeh, Sungbok Lee, Ulrich Neumann and Shrikanth Narayanan. 2004. Analysis of Emotion Recognition Using Facial Expressions, Speech and Multimodal Information. In *Proceedings of the 6th international conference on Multimodal interfaces*. pp. 205–211.

Campagna, Rachel L, Alexandra A Mislin, Dejun Tony Kong and William P Bottom. 2016. "Strategic Consequences of Emotional Misrepresentation in Negotiation: The Blowback Effect." *Journal of Applied Psychology* 101(5):605.

Carlsmith, Kevin M, John M Darley and Paul H Robinson. 2002. "Why Do We Punish? Deterrence and Just Deserts as Motives for Punishment." *Journal of Personality and Social Psychology* 83(2):284.

Carothers, Thomas and Andrew O'Donohue. 2019. *Democracies Divided: The Global Challenge of Political Polarization*. Brookings Institution Press.

Carter, David B and Curtis S Signorino. 2010. "Back to the Future: Modeling Time Dependence in Binary Data." *Political Analysis* 18(3):271–292.

Chaudoin, Stephen. 2014. "Promises or Policies? An Experimental Analysis of International Agreements and Audience Reactions." *International Organization* 68(1):235–256.

Cheshin, Arik, Adi Amit and Gerben A Van Kleef. 2018. "The Interpersonal Effects of Emotion Intensity in Customer Service: Perceived Appropriateness and Authenticity of Attendants' Emotional Displays Shape Customer Trust and Satisfaction." *Organizational Behavior and Human Decision Processes* 144:97–111.

Chiba, Daina, Jesse C Johnson and Brett Ashley Leeds. 2015. "Careful commitments: Democratic states and alliance design." *The Journal of Politics* 77(4):968–982.

Clifford, Scott, Geoffrey Sheagley and Spencer Piston. 2021. "Increasing Precision without Altering Treatment Effects: Repeated Measures Designs in Survey Experiments." *American Political Science Review* 115(3):1048–1065.

Coppedge, Michael, John Gerring, Carl Henrik Knutsen, Staffan I. Lindberg, Jan Teorell, David Altman, Michael Bernhard, Agnes Cornell, M. Steven Fish, Lisa Gastaldi, Haakon Gjerløw, Adam Glynn, Ana Good God, Sandra Grahn, Allen Hicken, Katrin Kinzelbach, Joshua Krusell, Kyle L. Marquardt, Kelly McMann, Valeriya Mechkova, Juraj Medzihorsky, Pamela Paxton, Daniel Pemstein, Josefine Pernes, Oskar Rydén, Johannes von Römer, Brigitte Seim, Rachel Sigman, Svend-Erik Skaaning, Jeffrey Staton, Aksel Sundström, Eitan Tzelgov, Yi-ting Wang, Tore Wig, Steven Wilson and Daniel Ziblatt. 2023. "V-Dem Country-Year Dataset v13.".

Coppock, Alexander. 2019. "Generalizing from Survey Experiments Conducted on Mechanical Turk: A Replication Approach." *Political Science Research and Methods* 7(3):613–628.

Côté, Stéphane, Ivona Hideg and Gerben A Van Kleef. 2013. "The consequences of faking anger in negotiations." *Journal of Experimental Social Psychology* 49(3):453–463.

Croco, Sarah E and Jessica LP Weeks. 2016. "War Outcomes and Leader Tenure." *World Politics* 68(4):577–607.

Croco, Sarah E, Michael J Hanmer and Jared A McDonald. 2021. "At What Cost? Reexamining Audience Costs in Realistic Settings." *The Journal of Politics* 83(1):8–22.

Dafoe, Allan, Baobao Zhang and Devin Caughey. 2018. "Information Equivalence in Survey Experiments." *Political Analysis* 26(4):399–416.

Dafoe, Allan, Jonathan Renshon and Paul Huth. 2014. "Reputation and Status as Motives for War." *Annual Review of Political Science* 17:371–393.

Dafoe, Allan, Sophia Hatz and Baobao Zhang. 2021. "Coercion and provocation." *Journal of Conflict Resolution* 65(2-3):372–402.

Darwin, Charles. 1872. *The Expression of the Emotions in Man and Animals*. University of Chicago Press.

De Melo, Celso M, Peter J Carnevale, Stephen J Read and Jonathan Gratch. 2014. "Reading People's Minds from Emotion Expressions in Interdependent Decision Making." *Journal of personality and social psychology* 106(1):73.

Deffenbacher, Jerry L, Eugene R Oetting, Rebekah S Lynch and Chad D Morris. 1996. "The Expression of Anger and Its Consequences." *Behaviour Research and Therapy* 34(7):575–590.

Downes, Alexander B and Todd S Sechser. 2012. "The illusion of democratic credibility." *International Organization* 66(3):457–489.

Duelfer, Charles. 2005. *Comprehensive Report of the Special Advisor to the DCI on Iraq's WMD, with Addendums*. Central Intelligence Agency.

Egami, Naoki and Erin Hartman. 2023. "Elements of External Validity: Framework, Design, and Analysis." *American Political Science Review* 117(3):1070–1088.

Ekman, Paul. 1992*a*. "An argument for basic emotions." *Cognition & emotion* 6(3-4):169–200.

Ekman, Paul. 1992*b*. *Telling Lies: Clues to Deceit in the Marketplace, Politics, and Marriage*. W.W. Norton & Company.

Ekman, Paul and Wallace V Friesen. 1971. "Constants across Cultures in the Face and Emotion." *Journal of Personality and Social Psychology* 17(2):124.

Ekman, Paul and Wallace V Friesen. 1974. "Detecting Deception from the Body or Face." *Journal of Personality and Social Psychology* 29(3):288.

Fearon, James D. 1994. "Domestic Political Audiences and the Escalation of International Disputes." *American Political Science Review* 88(3):577–592.

Fearon, James D. 1995. "Rationalist Explanations for War." *International organization* 49(3):379–414.

Fearon, James D. 1997. "Signaling Foreign Policy Interests: Tying Hands Versus Sinking Costs." *Journal of Conflict Resolution* 41(1):68–90.

Ferrer, Rebecca A and Erin M Ellis. 2019. "Moving beyond Categorization to Understand Affective Influences on Real World Health Decisions." *Social and Personality Psychology Compass* 13(11):e12502.

Filipowicz, Allan, Sigal Barsade and Shimul Melwani. 2011. "Understanding Emotional Transitions: the Interpersonal Consequences of Changing emotions in Negotiations." *Journal of personality and social psychology* 101(3):541.

Filson, Darren and Suzanne Werner. 2002. "A Bargaining Model of War and Peace: Anticipating the Onset, Duration, and Outcome of War." *American Journal of Political Science* 46(4):819–837.

Frank, Robert H. 1988. *Passions Within Reason: The Strategic Role of the Emotions.* WW Norton & Co.

Frijda, Nico H. 1986. *The Emotions*. Cambridge University Press.

Frijda, Nico H. and Batja Mesquita. 1994. The Social Roles and Functions of Emotions. In *Emotion and Culture: Empirical Studies of Mutual Influence*, ed. Shinobu Kitayama and Hazel R. Markus. American Psychological Association pp. 51–87.

Gartner, Scott Sigmund. 2008. "The Multiple Effects of Casualties on Public Support for War: An Experimental Approach." *American Political Science Review* 102(1):95–106.

Geddes, Barbara, Joseph Wright and Erica Frantz. 2014. "Autocratic Breakdown and Regime Transitions: A New Data Set." *Perspectives on Politics* 12(2):313–331.

Gelpi, Christopher F and Michael Griesdorf. 2001. "Winners or Losers? Democracies in International Crisis, 1918–94." *American Political Science Review* 95(3):633–647.

Gelpi, Christopher and Joseph M Grieco. 2015. "Competency Costs in Foreign Affairs: Presidential Performance in International Conflicts and Domestic Legislative Success, 1953–2001." *American Journal of Political Science* 59(2):440–456.

Gelpi, Christopher, Peter D Feaver and Jason Reifler. 2009. *Paying the Human Costs of War: American Public Opinion and Casualties in Military Conflicts*. Princeton University Press.

George, Alexander L. 1991. *Forceful Persuasion: Coercive Diplomacy as an Alternative to War*. US Institute of Peace Press.

Goemans, Henk E, Kristian Skrede Gleditsch and Giacomo Chiozza. 2009. "Introducing Archigos: A Dataset of Political Leaders." *Journal of Peace Research* 46(2):269–283.

Grimmer, Justin and Brandon M Stewart. 2013. "Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts." *Political analysis* 21(3):267–297.

Guisinger, Alexandra and Alastair Smith. 2002. "Honest Threats: The Interaction of Reputation and Political Institutions in International Crises." *Journal of Conflict Resolution* 46(2):175–200.

Hafner-Burton, Emilie M, Stephan Haggard, David A Lake and David G Victor. 2017. "The behavioral revolution and international relations." *International Organization* 71(S1):S1–S31.

Hall, Todd H. 2015. *Emotional Diplomacy: Official Emotion on the International Stage*. Cornell University Press.

Hall, Todd H. 2017. "On Provocation: Outrage, International Relations, and the Franco–Prussian War." *Security Studies* 26(1):1–29.

Hall, Todd H and Andrew AG Ross. 2015. "Affective Politics after 9/11." *International Organization* 69(4):847–879.

Hall, Todd H and Keren Yarhi-Milo. 2012. "The Personal Touch: Leaders' Impressions, Costly Signaling, and Assessments of Sincerity in International Affairs." *International Studies Quarterly* 56(3):560–573.

Halperin, Eran, Daphna Canetti-Nisim and Sivan Hirsch-Hoefler. 2009. "The central role of group-based hatred as an emotional antecedent of political intolerance: Evidence from Israel." *Political Psychology* 30(1):93–123.

Hareli, Shlomo, Raveh Harush, Ramzi Suleiman, Michel Cossette, Stephanie Bergeron, Veronique Lavoie, Guillaume Dugay and Ursula Hess. 2009. "When scowling may be a good thing: The influence of anger expressions on credibility." *European Journal of Social Psychology* 39(4):631–638.

Hareli, Shlomo and Ursula Hess. 2010. "What Emotional Reactions Can Tell Us About the Nature of Others: An Appraisal Perspective on Person Perception." *Cognition and Emotion* 24(1):128–140.

Harmon-Jones, Eddie and John JB Allen. 1998. "Anger and Frontal Brain Activity: EEG Asymmetry Consistent with Approach Motivation Despite Negative Affective Valence." *Journal of Personality and Social Psychology* 74(5):1310.

Harris, Benjamin Norwood and Erik Lin-Greenberg. 2024. "Cheap Tweets?: Crisis Signaling in the Age of Twitter." *International Studies Quarterly* 68(2):sqae067.

Harvey, Frank P and John Mitton. 2016. *Fighting for Credibility: US Reputation and International Politics*. University of Toronto Press.

Herrmann, Richard K. 2017. "How attachments to the nation shape beliefs about the world: A theory of motivated reasoning." *International Organization* 71(S1):S61–S84.

Herrmann, Richard K, Philip E Tetlock and Penny S Visser. 1999. "Mass Public Decisions on Go to War: A Cognitive-Interactionist Framework." *American Political Science Review* 93(3):553–573.

Hideg, Ivona and Gerben A van Kleef. 2017. "When Expressions of Fake Emotions Elicit Negative Reactions: The Role of Observers' Dialectical Thinking." *Journal of Organizational Behavior* 38(8):1196–1212.

Holmes, Marcus. 2013. "The Force of Face-to-Face Diplomacy: Mirror Neurons and the Problem of Intentions." *International Organization* 67(4):829–861.

Hopf, Ted. 1994. *Peripheral visions: Deterrence theory and American foreign policy in the third world, 1965-1990*. University of Michigan Press.

Hupka, Ralph B, Alison P Lenton and Keith A Hutchison. 1999. "Universal Development of Emotion Categories in Natural Language." *Journal of Personality and Social Psychology* 77(2):247.

Hutcherson, Cendri A and James J Gross. 2011. "The Moral Emotions: A Social-Functionalist Account of Anger, Disgust, and Contempt." *Journal of Personality and Social Psychology* 100(4):719.

Huth, Paul K. 1997. "Reputations and Deterrence: A Theoretical and Empirical Assessment." *Security Studies* 7(1):72–99.

Imai, Kosuke, Luke Keele and Teppei Yamamoto. 2010. "Identification, Inference and Sensitivity Analysis for Causal Mediation Effects." *Statistical Science* 25(1):51–71.

Imai, Kosuke and Teppei Yamamoto. 2013. "Identification and Sensitivity Analysis for Multiple Causal Mechanisms: Revisiting Evidence from Framing Experiments." *Political Analysis* 21(2):141–171.

Jervis, Robert. 1970. *The Logic of Images in International Relations*. Columbia University Press.

Jervis, Robert. 1976. *Perception and Misperception in International Politics*. Princeton University Press.

Jervis, Robert. 2017. *How Statesmen Think: The Psychology of International Politics*. Princeton University Press.

Katagiri, Azusa and Eric Min. 2019. "The Credibility of Public and Private Signals: A Document-Based Approach." *American Political Science Review* 113(1):156–172.

Keltner, Dacher and Jonathan Haidt. 1999. "Social Functions of Emotions at Four Levels of Analysis." *Cognition & Emotion* 13(5):505–521.

Kertzer, Joshua. 2016. *Resolve in International Politics*. Vol. 2 Princeton University Press.

Kertzer, Joshua D. 2022. "Re-Assessing Elite-Public Gaps in Political Behavior." *American Journal of Political Science* 66(3):539–553.

Kertzer, Joshua D, Brian C Rathbun and Nina Srinivasan Rathbun. 2020. "The Price of Peace: Motivated Reasoning and Costly Signaling in International Relations." *International Organization* 74(1):95–118.

Kertzer, Joshua D and Dustin Tingley. 2018. "Political psychology in international relations: beyond the paradigms." *Annual Review of Political Science* 21:319–339.

Kertzer, Joshua D, Jonathan Renshon and Keren Yarhi-Milo. 2021. "How Do Observers Assess Resolve?" *British Journal of Political Science* 51(1):308–330.

Kertzer, Joshua D, Kathleen E Powers, Brian C Rathbun and Ravi Iyer. 2014. "Moral support: How moral values shape foreign policy attitudes." *The Journal of Politics* 76(3):825–840.

Kertzer, Joshua D and Ryan Brutger. 2016. "Decomposing Audience Costs: Bringing the Audience Back into Audience Cost Theory." *American Journal of Political Science* 60(1):234–249.

Koschut, Simon, Todd H Hall, Reinhard Wolf, Ty Solomon, Emma Hutchison and Roland Bleiker. 2017. "Discourse and Emotions in International Relations." *International Studies Review* 19(3):481–508.

Kurizaki, Shuhei. 2007. "Efficient Secrecy: Public versus Private Threats in Crisis Diplomacy." *American Political Science Review* 101(3):543–558.

Laukka, Petri and Hillary Anger Elfenbein. 2021. "Cross-Cultural Emotion Recognition and In-Group Advantage in Vocal Expression: A Meta-Analysis." *Emotion Review* 13(1):3–11.

Lazarus, Richard S. 1991. "Cognition and Motivation in Emotion." *American Psychologist* 46(4):352.

Leetaru, Kalev. 2010. "The Scope of FBIS and BBC Open-Source Media Coverage (U), 1979–2008." *Studies in Intelligence* 54(1):17–37.

Lelieveld, Gert-Jan, Eric Van Dijk, Ilja Van Beest and Gerben A Van Kleef. 2012. "Why Anger and Disappointment Affect Other's Bargaining Behavior Differently: The Moderating Role of Power and the Mediating Role of Reciprocal and Complementary Emotions." *Personality and Social Psychology Bulletin* 38(9):1209–1221.

Lelieveld, Gert-Jan, Eric Van Dijk, Ilja Van Beest, Wolfgang Steinel and Gerben A Van Kleef. 2011. "Disappointed in you, angry about your offer: Distinct negative emotions induce concessions via different mechanisms." *Journal of Experimental Social Psychology* 47(3):635–641.

Lerner, Jennifer S and Dacher Keltner. 2001. "Fear, Anger, and Risk." *Journal of Personality and Social Psychology* 81(1):146.

Lerner, Jennifer S and Larissa Z Tiedens. 2006. "Portrait of the Angry Decision Maker: How Appraisal Tendencies Shape Anger's Influence on Cognition." *Journal of Behavioral Decision Making* 19(2):115–137.

Levendusky, Matthew S and Michael C Horowitz. 2012. "When backing down is the right decision: Partisanship, new information, and audience costs." *The Journal of Politics* 74(2):323–338.

Lin-Greenberg, Erik. 2019. "Backing up, not backing down: Mitigating audience costs through policy substitution." *Journal of Peace Research* 56(4):559–574.

Lupton, Danielle L. 2020. *Reputation for Resolve: How Leaders Signal Determination in International Politics*. Cornell University Press.

Magaloni, Beatriz, Jonathan Chu and Eric Min. 2013. "Autocracies of the World, 1950-2012 (Version 1.0).". Dataset, Stanford University.

Maier, Jürgen and Alessandro Nai. 2020. "Roaring Candidates in the Spotlight: Campaign Negativity, Emotions, and Media Coverage in 107 National Elections." *The International Journal of Press/Politics* 25(4):576–606.

Malatesta, Carol Zander and Jeannette M Haviland. 1982. "Learning Display Rules: The Socialization of Emotion Expression in Infancy." *Child development* 53(4):991–1003.

Marshall, Monty and Keith Jaggers. 2002. "Polity IV Project: Political Regime Characteristics and Transitions, 1800–2002." *University of Maryland, College Park* .

Masterson, Michael. 2022. "Humiliation and International Conflict Preferences." *The Journal of Politics* 84(2):874–888.

Matsumoto, David. 1990. "Cultural similarities and differences in display rules." *Motivation and emotion* 14(3):195–214.

Matsumoto, David, Seung Hee Yoo and Johnny Fontaine. 2008. "Mapping expressive differences around the world: The relationship between emotional display rules and individualism versus collectivism." *Journal of cross-cultural psychology* 39(1):55–74.

Mattes, Michaela and Jessica LP Weeks. 2019. "Hawks, Doves, and Peace: An Experimental Approach." *American Journal of Political Science* 63(1):53–66.

Mattes, Michaela and Jessica LP Weeks. 2022. "Reacting to the Olive Branch: Hawks, Doves, and Public Support for Cooperation." *International Organization* 76(4):957–976.

Mayer, John D, David R Caruso and Peter Salovey. 1999. "Emotional Intelligence Meets Traditional Standards for an Intelligence." *Intelligence* 27(4):267–298.

McDermott, Rose. 2001. *Risk-Taking in International Politics: Prospect Theory in American Foreign Policy*. University of Michigan Press.

McManus, Roseanne W. 2014. "Fighting words: The effectiveness of statements of resolve in international conflict." *Journal of Peace Research* 51(6):726–740.

McManus, Roseanne W. 2017. *Statements of Resolve: Achieving Coercive Credibility in International Conflict*. Cambridge University Press.

Mehu, Marc, Marcello Mortillaro, Tanja Bänziger and Klaus R Scherer. 2012. "Reliable Facial Muscle Activation Enhances Recognizability and Credibility of Emotional Expression." *Emotion* 12(4):701.

Meikle, James and Ian Black. 2011. "Muammar Gaddafi Says He Will Die a Martyr Rather than Quit." *The Guardian* .
URL: *https://www.theguardian.com/world/2011/feb/22/muammar-gaddafi-libyan-tv-martyr*

Memmott, Mark. 2011. "Flashback: Reagan Calls Gadhafi The 'Mad Dog Of The Middle East." *NPR* .
URL: *https://www.npr.org/sections/thetwo-way/2011/02/22/133970620/flashback-reagan-calls-gadhafi-the-mad-dog-of-the-middle-east*

Mercer, Jonathan. 1996. *Reputation and international politics*. Cornell University Press.

Mercer, Jonathan. 2010. "Emotional Beliefs." *International Organization* 64(1):1–31.

Mercer, Jonathan. 2014. "Feeling like a State: Social Emotion and Identity." *International Theory* 6(3):515–535.

Miller, Steven V. 2021. *peacesciencer: A User's Guide for Quantitative Peace Science in R*.
URL: *http://svmiller.com/peacesciencer*

Modell, John and Timothy Haggerty. 1991. "The Social Impact of War." *Annual Review of Sociology* 17(1):205–224.

Mohammad, Saif M and Peter D Turney. 2013. "NRC Emotion Lexicon." *National Research Council, Canada* 2:234.

Monroy, M. C., A. S. Cowen and D. Keltner. 2022. "Intersectionality in Emotion Signaling and Recognition: The Influence of Gender, Ethnicity, and Social Class." *Emotion* 22(8):1980–1988.

Montgomery, Jacob M, Brendan Nyhan and Michelle Torres. 2018. "How conditioning on posttreatment variables can ruin your experiment and what to do about It." *American Journal of Political Science* 62(3):760–775.

Morrow, James D. 1989. "Capabilities, Uncertainty, and Resolve: A Limited Information Model of Crisis Bargaining." *American Journal of Political Science* 33(4):941–972.

Mufson, Steven. 2015. "Bush Saw Putin's 'Soul.' Obama Wants to Appeal to His Brain." *The Washington Post* .
URL: *https://www.washingtonpost.com/business/economy/bush-saw-putins-soul-obama-wants-to-appeal-to-his-brain/2015/12/01/264f0c7c-984b-11e5-8917-653b65c809eb$_s$tory.html*

Nomikos, William G and Nicholas Sambanis. 2019. "What Is the Mechanism Underlying Audience Costs? Incompetence, Belligerence, and Inconsistency." *Journal of Peace Research* 56(4):575–588.

Palmer, Glenn, Roseanne W McManus, Vito D'Orazio, Michael R Kenwick, Mikaela Karstens, Chase Bloch, Nick Dietrich, Kayla Kahn, Kellan Ritter and Michael J Soules. 2022. "The MID5 Dataset, 2011–2014: Procedures, Coding Rules, and Description." *Conflict Management and Peace Science* 39(4):470–482.

Petkanopoulou, Katerina, Rosa Rodríguez-Bailón, Guillermo B Willis and Gerben A van Kleef. 2019. "Powerless people don't yell but tell: The effects of social power on direct and indirect expression of anger." *European Journal of Social Psychology* 49(3):533–547.

Pillutla, Madan M and J Keith Murnighan. 1996. "Unfairness, Anger, and Spite: Emotional Rejections of Ultimatum Offers." *Organizational Behavior and Human Decision Processes* 68(3):208–224.

Plant, E Ashby, Janet Shibley Hyde, Dacher Keltner and Patricia G Devine. 2000. "The Gender Stereotyping of Emotions." *Psychology of Women Quarterly* 24(1):81–92.

Plutchik, Robert. 2001. "The Nature of Emotions: Human Emotions Have Deep Evolutionary Roots, a Fact That May Explain Their Complexity and Provide Tools for Clinical Practice." *American Scientist* 89(4):344–350.

Porter, Stephen, Leanne Ten Brinke and Brendan Wallace. 2012. "Secrets and Lies: Involuntary Leakage in Deceptive Facial Expressions as a Function of Emotional Intensity." *Journal of Nonverbal Behavior* 36:23–37.

Potter, Philip BK and Matthew A Baum. 2010. "Democratic Peace, Domestic Audience costs, and Political Communication." *Political Communication* 27(4):453–470.

Potter, Philip BK and Matthew A Baum. 2014. "Looking for Audience Costs in All the Wrong Places: Electoral Institutions, Media Access, and Democratic Constraint." *The Journal of Politics* 76(1):167–181.

Powers, Kathleen E and Dan Altman. 2023. "The Psychology of Coercion Failure: How Reactance Explains Resistance to Threats." *American Journal of Political Science* 67(1):221–238.

Powers, Ryan and Jonathan Renshon. 2021. "International Status Concerns and Domestic Support for Political Leaders." *American Journal of Political Science* 67(3):732–747.

Press, Daryl Grayson. 2005. *Calculating Credibility: How Leaders Assess Military Threats*. Cornell University Press.

Quek, Kai. 2016. "Are Costly Signals More Credible? Evidence of Sender-Receiver Gaps." *The Journal of Politics* 78(3):925–940.

Quigley, Brian M and James T Tedeschi. 1996. "Mediating Effects of Blame Attributions on Feelings of Anger." *Personality and Social Psychology Bulletin* 22(12):1280–1288.

Ramsay, Kristopher W. 2017. "Information, Uncertainty, and War." *Annual Review of Political Science* 20:505–527.

Rawtani, Deepak, Gunjan Gupta, Nitasha Khatri, Piyush K Rao and Chaudhery Mustansar Hussain. 2022. "Environmental Damages due to War in Ukraine: A Perspective." *Science of The Total Environment* 850:157932.

Reed, Lawrence Ian, Peter DeScioli and Steven A Pinker. 2014. "The Commitment Function of Angry Facial Expressions." *Psychological Science* 25(8):1511–1517.

Renshon, Jonathan, Allan Dafoe and Paul Huth. 2018. "Leader Influence and Reputation Formation in World Politics." *American Journal of Political Science* 62(2):325–339.

Renshon, Jonathan, Julia J Lee and Dustin Tingley. 2017. "Emotions and the Micro-Foundations of Commitment Problems." *International Organization* 71(S1):S189–S218.

Roop, Joseph E. 1969. *Foreign Broadcast Information Service History, Part I, 1941-1947*. Central Intelligence Agency.

Ryan, Timothy J. 2017. "No Compromise: Political Consequences of Moralized Attitudes." *American Journal of Political Science* 61(2):409–423.

Sanfey, Alan G, James K Rilling, Jessica A Aronson, Leigh E Nystrom and Jonathan D Cohen. 2003. "The Neural Basis of Economic Decision-Making in the Ultimatum Game." *Science* 300(5626):1755–1758.

Sartori, Anne E. 2002. "The Might of the Pen: A Reputational Theory of Communication in International Disputes." *International Organization* 56(1):121–149.

Sartori, Anne E. 2005. *Deterrence by Diplomacy*. Princeton University Press.

Sasley, Brent E. 2011. "Theorizing states' emotions." *International Studies Review* 13(3):452–476.

Schelling, Thomas C. 1960. *The Strategy of Conflict*. Harvard university press.

Schelling, Thomas C. 1966. *Arms and Influence*. Yale University Press.

Schub, Robert. 2022. "Informing the leader: Bureaucracies and international crises." *American Political Science Review* 116(4):1460–1476.

Schultz, Kenneth A. 1998. "Domestic Opposition and Signaling in International Crises." *American Political Science Review* 92(4):829–844.

Schultz, Kenneth A. 2001. *Democracy and Coercive Diplomacy*. Cambridge University Press.

Sechser, Todd S. 2011. "Militarized Compellent Threats, 1918–2001." *Conflict Management and Peace Science* 28(4):377–401.

Sell, Aaron, Daniel Sznycer, Laith Al-Shawaf, Julian Lim, Andre Krauss, Aneta Feldman, Ruxandra Rascanu, Lawrence Sugiyama, Leda Cosmides and John Tooby. 2017. "The Grammar of Anger: Mapping the Computational Architecture of a Recalibrational Emotion." *Cognition* 168:110–128.

Sell, Aaron, John Tooby and Leda Cosmides. 2009. "Formidability and the Logic of Human Anger." *Proceedings of the National Academy of Sciences* 106(35):15073–15078.

Sell, Aaron N. 2011. "The Recalibrational Theory and Violent Anger." *Aggression and Violent Behavior* 16(5):381–389.

Shelef, Nadav G. 2016. "Unequal ground: Homelands and conflict." *International Organization* 70(1):33–63.

Singer, J David. 1988. "Reconstructing the Correlates of War Dataset on Material Capabilities of States, 1816–1985." *International Interactions* 14(2):115–132.

Slantchev, Branislav L. 2003. "The principle of convergence in wartime negotiations." *American Political Science Review* 97(4):621–632.

Slantchev, Branislav L. 2005. "Military Coercion in Interstate Crises." *American Political Science Review* 99(4):533–547.

Smith, Alastair. 1998. "International crises and domestic politics." *American Political Science Review* 92(3):623–638.

Smith, Bradley C and William Spaniel. 2019. "Militarized Disputes, Uncertainty, and Leader Tenure." *Journal of Conflict Resolution* 63(5):1222–1252.

Smith, Ron P. 2014. "The Economic Costs of Military Conflict." *Journal of Peace Research* 51(2):245–256.

Snyder, Glenn Herald. 1961. "Deterrence and Defense.".

Snyder, Glenn Herald and Paul Diesing. 1977. *Conflict Among Nations: Bargaining, Decision Making, and System Structure in International Crises*. Princeton University Press.

Snyder, Jack and Erica D Borghard. 2011. "The Cost of Empty Threats: A Penny, Not a Pound." *American Political Science Review* 105(3):437–456.

Spielberger, Charles D, Sumner J Sydeman, Ashley E Owen and Brian J Marsh. 1999. *Measuring Anxiety and Anger with the State-Trait Anxiety Inventory (STAI) and the State-Trait Anger Expression Inventory (STAXI)*. Lawrence Erlbaum Associates Publishers.

Szczurek, Lauren, Benoît Monin and James J Gross. 2012. "The Stranger Effect: The Rejection of Affective Deviants." *Psychological science* 23(10):1105–1111.

Sznycer, Daniel, Aaron Sell and Alexandre Dumont. 2022. "How Anger Works." *Evolution and Human Behavior* 43(2):122–132.

Tetlock, Philip E. 2003. "Thinking the Unthinkable: Sacred Values and Taboo Cognitions." *Trends in Cognitive Sciences* 7(7):320–324.

Thompson, William and David Dreyer. 2011. *Handbook of international rivalries*. CQ Press.

Tingley, Dustin, Teppei Yamamoto, Kentaro Hirose, Luke Keele and Kosuke Imai. 2014. "mediation: R Package for Causal Mediation Analysis." *Journal of Statistical Software* 59(5):1–38.

Tng, Han-Ying and Al KC Au. 2014. "Strategic Display of Anger and Happiness in Negotiation: The Moderating Role of Perceived Authenticity." *Negotiation Journal* 30(3):301–327.

Tomz, Michael. 2007*a*. "Domestic audience costs in international relations: An experimental approach." *International Organization* 61(4):821–840.

Tomz, Michael. 2007*b*. *Reputation and International Cooperation: Sovereign Debt across Three Centuries*. Princeton University Press.

Trachtenberg, Marc. 2012. "Audience Costs: An Historical Analysis." *Security Studies* 21(1):3–42.

Tracy, Jessica L and Richard W Robins. 2008. "The Automaticity of Emotion Recognition." *Emotion* 8(1):81.

Trager, Robert F. 2010. "Diplomatic Calculus in Anarchy: How Communication Matters." *American Political Science Review* 104(2):347–368.

Trager, Robert F. 2017. *Diplomacy: Communication and the origins of international order*. Cambridge University Press.

Trager, Robert F and Lynn Vavreck. 2011. "The Political Costs of Crisis Bargaining: Presidential Rhetoric and the Role of Party." *American Journal of Political Science* 55(3):526–545.

Van Beest, Ilja, Gerben A Van Kleef and Eric Van Dijk. 2008. "Get Angry, Get Out: The Interpersonal Effects Of Anger Communication in Multiparty Negotiation." *Journal of Experimental Social Psychology* 44(4):993–1002.

Van Doorn, Evert A, Marc W Heerdink and Gerben A Van Kleef. 2012. "Emotion and the Construal of Social Situations: Inferences of Cooperation Versus Competition from Expressions of Anger, Happiness, and Disappointment." *Cognition & Emotion* 26(3):442–461.

Van Kleef, Gerben A. 2009. "How emotions regulate social life: The emotions as social information (EASI) model." *Current directions in psychological science* 18(3):184–188.

Van Kleef, Gerben A, Carsten KW De Dreu and Antony SR Manstead. 2004. "The Interpersonal Effects of Anger and Happiness in Negotiations." *Journal of Personality and Social Psychology* 86(1):57.

Van Kleef, Gerben A, Carsten KW De Dreu and Antony SR Manstead. 2010. An Interpersonal Approach to Emotion in Social Decision Making: The Emotions as Social Information Model. In *Advances in Experimental Social Psychology*. Vol. 42 Elsevier pp. 45–96.

Van Kleef, Gerben A, Carsten KW De Dreu, Davide Pietroni and Antony SR Manstead. 2006. "Power and Emotion in Negotiation: Power Moderates the Interpersonal Effects of Anger and Happiness on Concession Making." *European Journal of Social Psychology* 36(4):557–581.

Van Leeuwen, Boris, Charles N Noussair, Theo Offerman, Sigrid Suetens, Matthijs Van Veelen and Jeroen Van De Ven. 2018. "Predictably Angry—Facial Cues Provide a Credible Signal of Destructive Behavior." *Management Science* 64(7):3352–3364.

Veenstra, Lotte, Brad J Bushman and Sander L Koole. 2018. "The Facts on the Furious: A Brief Review of the Psychology of Trait Anger." *Current Opinion in Psychology* 19:98–103.

Walter, Barbara F. 2006. "Building Reputation: Why Governments Fight Some Separatists but Not Others." *American Journal of Political Science* 50(2):313–330.

Watson, David and Lee Anna Clark. 1994. *The PANAS-X: Manual for the Positive and Negative Affect Schedule–Expanded Form*. University of Iowa.

Weeks, Jessica L. 2008. "Autocratic Audience Costs: Regime Type and Signaling Resolve." *International Organization* 62(1):35–64.

Weeks, Jessica LP. 2014. *Dictators at war and peace*. Cornell University Press.

Weisiger, Alex and Keren Yarhi-Milo. 2015. "Revisiting Reputation: How Past Actions Matter in International Politics." *International Organization* 69(2):473–495.

Wolford, Scott. 2007. "The Turnover Trap: New Leaders, Reputation, and International Conflict." *American Journal of Political Science* 51(4):772–788.

Wong, Seanon S. 2016. "Emotions and the Communication of Intentions in Face-to-Face Diplomacy." *European Journal of International Relations* 22(1):144–167.

Wong, Seanon S. 2019. "Stoics and Hotheads: Leaders' Temperament, Anger, and the Expression of Resolve in Face-to-Face Diplomacy." *Journal of Global Security Studies* 4(2):190–208.

Wood, B Dan. 2012. *Presidential Saber Rattling: Causes and Consequences*. Cambridge University Press.

Woolley, John T and Gerhard Peters. 1999. "The American Presidency Project." *Santa Barbara, CA. Available from http://www.presidency.ucsb.edu/ws* .

Yamagishi, Toshio, Yutaka Horita, Haruto Takagishi, Mizuho Shinada, Shigehito Tanida and Karen S Cook. 2009. "The Private Rejection of Unfair Offers and Emotional Commitment." *Proceedings of the National Academy of Sciences* 106(28):11520–11523.

Yarhi-Milo, Keren. 2013. "In the Eye of the Beholder: How Leaders and Intelligence Communities Assess the Intentions of Adversaries." *International Security* 38(1):7–51.

Yarhi-Milo, Keren. 2014. *Knowing the adversary: leaders, intelligence, and assessment of intentions in international relations*. Vol. 146 Princeton University Press.

Yarhi-Milo, Keren. 2018. *Who Fights for Reputation: The Psychology of Leaders in International Conflict*. Vol. 156 Princeton University Press.

Yarhi-Milo, Keren, Joshua D Kertzer and Jonathan Renshon. 2018. "Tying Hands, Sinking Costs, and Leader Attributes." *Journal of Conflict Resolution* 62(10):2150–2179.

Zeitzoff, Thomas. 2014. "Anger, Exposure to Violence, and Intragroup Conflict: A "Lab in the Field" Experiment in Southern Israel." *Political Psychology* 35(3):309–335.